# Stability and Uniform Approximation of Nonlinear Filters using the Hilbert Metric, and Application to Particle Filters

François Le Gland, Nadia Oudjane

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Stability and Uniform Approximation of Nonlinear Filters using the Hilbert Metric, and Application to Particle Filters*

François Le Gland, Nadia Oudjane

**N˚4215**

Juin 2001

THÈME 4

ISSN 0249-6399 ISRN INRIA/RR--4215--FR+ENG

Rapport de recherche

# Stability and Uniform Approximation of Nonlinear Filters using the Hilbert Metric, and Application to Particle Filters

François Le Gland*, Nadia Oudjane**

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet SIGMA2 ($\sigma^2$)

Rapport de recherche n ° 4215 — Juin 2001 — 30 pages

**Abstract:** We study the stability of the optimal filter w.r.t. its initial condition and w.r.t. the model for the hidden state and the observations in a general hidden Markov model, using the Hilbert projective metric. These stability results are then used to prove, under some mixing assumption, the uniform convergence to the optimal filter of several particle filters, such as the interacting particle filter and other original particle filters.

**Key-words:** hidden Markov model, nonlinear filter, particle filter, stability, Hilbert metric, total variation norm, mixing, regularizing kernel.

*(Résumé : tsvp)*

This work was partially supported by the CNRS programme *Modélisation et Simulation Numérique*.

* IRISA / INRIA, Campus de Beaulieu, 35042 RENNES Cédex, France. — legland@irisa.fr
** Département de Mathématiques, Université Paris XIII, 93430 VILLETANEUSE, France. — oudjane@maths.univ-paris13.fr

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : 02 99 84 71 00 - International : +33 2 99 84 71 00
Télécopie : 02 99 84 71 71 - International : +33 2 99 84 71 71

# Stabilité et approximation uniforme des filtres non–linéaires avec la métrique de Hilbert, et application aux filtres particulaires

**Résumé :** Nous étudions la stabilité du filtre optimal par rapport à sa condition initiale et par rapport au modèle décrivant l'état caché et les observations dans un modèle de Markov caché général, en utilisant la métrique de Hilbert. Ces résultats de stabilité sont ensuite utilisés pour démontrer, sous une hypothèse de mélange, la convergence uniforme vers le filtre optimal de plusieurs filtres particulaires, tels que le filtre particulaire avec interaction et d'autres filtres particulaires originaux.

**Mots-clé :** modèle de Markov caché, filtre non–linéaire, filtre particulaire, stabilité, métrique de Hilbert, norme en variation totale, mélange, noyau régularisant.

# 1 Introduction

The stability of the optimal filter has become recently an active research area. Kunita in [19] and Stettner in [28] have studied the ergodic properties of the measure–valued process formed by the optimal filter, and have proved the existence of a unique invariant probability distribution for this process. Ocone and Pardoux have proved in [24] that the filter forgets its initial condition in the $L^p$ sense, without stating any rate of convergence. Recently, a new approach has been proposed by Da Prato, Fuhrman and Malliavin in [8], using the Hilbert projective metric. This metric allows to get rid of the normalization constant in the Bayes formula, and reduces the problem to studying the linear equation satisfied by the unnormalized optimal filter. Using the Hilbert metric, stability results w.r.t. the initial condition have been proved by Atar and Zeitouni in [4], and some stability result w.r.t. the model have been proved by Le Gland and Mevel in [20, 21], for hidden Markov models (HMM) with finite state space. The results and methods of [4] have been extended to HMM with Polish state space by Atar and Zeitouni in [3]. Independently, Del Moral and Guionnet have adopted in [9], for the same class of HMM, another approach based on semi–group techniques and on the Dobrushin ergodic coefficient, to derive stability results w.r.t. the initial condition, which are used to prove uniform convergence of the interacting particle filter (IPF) to the optimal filter, with a rate $(1/\sqrt{N})^\alpha$ for any $\alpha < 1$. New approaches have been proposed recently, to prove the stability of the optimal filter w.r.t. its initial condition, in the case of a noncompact state space, see e.g. Atar [1], Atar, Viens and Zeitouni [2], Budhiraja and Ocone [6, 7].

In this article, we use the approach based on the Hilbert metric to study the asymptotic behavior of the optimal filter, and to prove as in [9] the uniform convergence of several particle filters, such as the IPF and original new particle filters.

A common assumption to prove stability results, see e.g. in [9], is that the Markov transition kernels are mixing, which implies that the hidden state sequence is ergodic. Our results are obtained under the assumption that the nonnegative kernels describing the evolution of the unnormalized optimal filter, and incorporating simultaneously the Markov transition kernels and the likelihood functions, are mixing. This is a weaker assumption, see Proposition 3.9, which allows to consider some cases, similar to the case studied in [6], where the hidden state sequence is not ergodic, see Example 3.10. This point of view will be further developped elsewhere. Our main contribution is to study also the stability of the optimal filter w.r.t. the model, when the local error is propagated by mixing kernels, and can be estimated in the Hilbert metric, in the total variation norm, or in a weaker distance suitable for random probability distributions.

Uniform convergence results of the IPF to the optimal filter are proved in [9] under an additional uniform lower bound assumption on the likelihood functions, which is rather strong. Our uniform convergence results are obtained under the assumption that the expected values of the likelihood functions integrated against any possible predicted probability distribution, are bounded away from zero. This assumption is automatically satisfied under our weaker mixing assumption, see Remark 5.6. Motivated by practical considerations, we introduce a variant of the IPF, where an adaptive number of particles is used, based on *a posteriori* estimates. The resulting sequential particle filter (SPF) is shown to converge uniformly to the optimal filter, independently of any lower bound assumption on the likelihood functions. The counterpart is that the computational time is random, and that the expected number of particles does depend on the integrated lower bounds of the likelihood functions. Also motivated by practical considerations, i.e. to avoid the *degeneracy of particle weights* and the *degeneracy of particle locations*, which are two known causes of divergence of particle filters, we introduce regularized particle filters (RPF), which are shown to converge uniformly to the optimal filter.

The paper is organized as follows : In the next section we define the framework of the nonlinear filtering problem and we introduce some notations. In Section 3, we state some properties of the Hilbert metric, which are used in Section 4 to prove the stability of the optimal filter w.r.t. its initial condition and w.r.t. the model. These stability results are used to prove the uniform convergence of several particle filters to the optimal filter. First, uniform convergence in the weak sense is proved in Section 5 for interacting particle filters, with a rate $1/\sqrt{N}$, and sequential particle filters, with a random number of particles, are also considered. Finally, regularized particle filters are defined in Section 6, for which uniform convergence in the weak sense and in the total variation norm are proved.

# 2 Optimal filter for general HMM

We consider the following model, with a hidden (non–observed) state sequence $\{X_n \, , \, n \geq 0\}$ and an observation sequence $\{Y_n \, , \, n \geq 1\}$, taking values in a complete separable metric space $E$ and in $F = \mathbb{R}^d$, respectively (in Section 6, it will be assumed that $E = \mathbb{R}^m$) :

- The state sequence $\{X_n \, , \, n \geq 0\}$ is defined as an inhomogeneous Markov chain, with transition probability kernel $Q_n$, i.e.

$$\mathbb{P}[X_n \in dx \mid X_{0:n-1} = x_{0:n-1}] = \mathbb{P}[X_n \in dx \mid X_{n-1} = x_{n-1}] = Q_n(x_{n-1}, dx) \ ,$$

  for all $n \geq 1$, and with initial probability distribution $\mu_0$. For instance, $\{X_n \, , \, n \geq 0\}$ could be defined by the following equation

$$X_n = f_n(X_{n-1}, W_n) \ ,$$

  where $\{W_n \, , \, n \geq 0\}$ is a sequence of independent random variables, not necessarily Gaussian, independent of the initial state $X_0$.

- The *memoryless channel* assumption holds, i.e. given the state sequence $\{X_n \, , \, n \geq 0\}$
  - the observations $\{Y_n \, , \, n \geq 1\}$ are independent random variables,
  - for all $n \geq 1$, the conditional probability distribution of $Y_n$ depends only on $X_n$.

  For instance, the observation sequence $\{Y_n \, , \, n \geq 1\}$ could be related to the state sequence $\{X_n \, , \, n \geq 0\}$ by

$$Y_n = h_n(X_n, V_n) \ ,$$

  for all $n \geq 1$, where $\{V_n \, , \, n \geq 1\}$ is a sequence of independent random variables, not necessarily Gaussian, independent of the state sequence $\{X_n \, , \, n \geq 0\}$. In addition, it is assumed that for all $n \geq 1$, the collection of probability distributions $\mathbb{P}[Y_n \in dy \mid X_n = x]$ on $F$, parametrized by $x \in E$, is dominated, i.e.

$$\mathbb{P}[Y_n \in dy \mid X_n = x] = g_n(x, y) \, \lambda_n^F(dy) \ ,$$

  for some nonnegative measure $\lambda_n^F$ on $F$. The corresponding likelihood function is defined by $\Psi_n(x) = g_n(x, Y_n)$, and depends implicitely on the observation $Y_n$.

The following notations and definitions will be used throughout the paper.

- The set of probability distributions on $E$, and the set of nonnegative measures on $E$, are denoted by $\mathcal{P}(E)$ and $\mathcal{M}^+(E)$ respectively.

- The notation $\| \cdot \|$ is used for the total variation norm on the set of signed measures on $E$, and for the supremum norm on the set of bounded measurable functions defined on $E$, depending on the context.

- With any nonnegative kernel $K$ defined on $E$, is associated a nonnegative linear operator acting on functions, denoted by $K^*$, and defined by

$$K^* \, \phi(x) = \int_E K(x, dx') \, \phi(x') \ ,$$

  for any measurable function $\phi$ defined on $E$. Consequently, the adjoint nonnegative linear operator acting on nonnegative measures, and denoted by $K$, is defined by

$$K \, \mu(dx') = \mu \, K^*(dx') = \int_E \mu(dx) \, K(x, dx') \ ,$$

  for any nonnegative measure $\mu$ on $E$.

- With any nonzero $\mu \in \mathcal{M}^+(E)$, i.e. such that $\mu(E) \neq 0$, is associated the normalized nonnegative measure (i.e. probability distribution) $\bar{\mu} = \mu / \mu(E) \in \mathcal{P}(E)$.

- With any nonnegative kernel $K$ defined on $E$, is associated the normalized nonnegative nonlinear operator $\bar{K}$ on $\mathcal{M}^+(E)$, taking values in $\mathcal{P}(E)$, and defined by $\bar{K}(\mu) = K \, \mu / (K \, \mu)(E) = K \, \bar{\mu} / (K \, \bar{\mu})(E) = \bar{K}(\bar{\mu})$ for any $\mu \in \mathcal{M}^+(E)$ such that $K \, \mu(E) \neq 0$, and by $\bar{K}(\mu) = 0$ otherwise.

The problem of nonlinear filtering is to compute at each time $n$, the conditional probability distribution $\mu_n$ of the state $X_n$ given the observation sequence $Y_{1:n} = (Y_1, \cdots, Y_n)$ up to time $n$. The transition from $\mu_{n-1}$ to $\mu_n$ is described by the following diagram

$$\mu_{n-1} \xrightarrow[\text{prediction}]{\qquad\qquad} \mu_{n|n-1} = Q_n\, \mu_{n-1} \xrightarrow[\text{correction}]{\qquad\qquad} \mu_n = \Psi_n \cdot \mu_{n|n-1} \ ,$$

where $\cdot$ denotes the projective product.

**Remark 2.1.** Notice that the normalizing constant $\langle \mu_{n|n-1}, \Psi_n \rangle$ is a.s. positive. Indeed

$$\begin{aligned} \mathbb{P}[Y_n \in dy \mid Y_{1:n-1}] &= \int_E \mathbb{P}[Y_n \in dy \mid X_n = x]\; \mathbb{P}[X_n \in dx \mid Y_{1:n-1}] \\ &= [\int_E g_n(x,y)\, \mu_{n|n-1}(dx)\,]\; \lambda_n^F(dy) = \ell_n(y)\, \lambda_n^F(dy) \ , \end{aligned}$$

hence

$$\langle \mu_{n|n-1}, \Psi_n \rangle = \int_E g_n(x, Y_n)\, \mu_{n|n-1}(dx) = \ell_n(Y_n) \ .$$

Therefore

$$\mathbb{P}[\, \langle \mu_{n|n-1}, \Psi_n \rangle = 0 \mid Y_{1:n-1}\,] \;=\; \int_F \mathbf{1}_{\{\ell_n(y) = 0\}}\; \ell_n(y)\, \lambda_n^F(dy) = 0 \ .$$

**Remark 2.2.** Notice also that, for any test function $\psi$ defined on $F$

$$\mathbb{E}[\, \frac{\psi(Y_n)}{\langle \mu_{n|n-1}, \Psi_n \rangle} \mid Y_{1:n-1}] = \mathbb{E}[\, \frac{\psi(Y_n)}{\ell_n(Y_n)} \mid Y_{1:n-1}] = \int_F \psi(y)\, \lambda_n^F(dy) \ .$$

In particular, if $\psi(y) = g_n(x,y)$, then $\psi(Y_n) = \Psi_n(x)$, and

$$\mathbb{E}[\, \frac{\Psi_n(x)}{\langle \mu_{n|n-1}, \Psi_n \rangle} \mid Y_{1:n-1}] = \int_F g_n(x,y)\, \lambda_n^F(dy) = 1 \ ,$$

for any $x \in E$.

For any $n \geq 1$, we introduce the nonnegative kernel

$$R_n(x, dx') = Q_n(x, dx')\, \Psi_n(x') \ ,$$

and the associated nonnegative linear operator $R_n = \Psi_n\, Q_n$ on $\mathcal{M}^+(E)$, defined by

$$R_n\, \mu(dx') = \int_E \mu(dx)\, Q_n(x, dx')\, \Psi_n(x') \ ,$$

for any $\mu \in \mathcal{M}^+(E)$. Notice that $R_n$ depends on the observation $Y_n$ through the likelihood function $\Psi_n$. With this definition, $(R_n\, \mu_{n-1})(E) = \langle \mu_{n|n-1}, \Psi_n \rangle$ is a.s. positive, the evolution of the optimal filter can be written as follows

$$\mu_n = \Psi_n \cdot (Q_n\, \mu_{n-1}) = \frac{R_n\, \mu_{n-1}}{(R_n\, \mu_{n-1})(E)} = \bar{R}_n(\mu_{n-1}) \ , \tag{1}$$

and iteration yields

$$\mu_n = \bar{R}_n(\mu_{n-1}) = \bar{R}_n \circ \cdots \circ \bar{R}_m(\mu_{m-1}) = \bar{R}_{n:m}(\mu_{m-1}) \ .$$

Equation (1) shows clearly that the evolution of the optimal filter is nonlinear only because of the normalization term coming from the Bayes rule. In the following section a projective metric is introduced precisely to get rid of the normalization and to come down to the analysis of a linear evolution.

# 3 Hilbert metric on the set of finite nonnegative measures

In this section we recall the definition of the Hilbert metric and its associated contraction coefficient, the Birkhoff contraction coefficient. We introduce also a mixing property for nonnegative kernels, and we state some properties relating the Hilbert metric with other distances on the set of probability distributions, e.g. the total variation norm, or a weaker distance suitable for random probability distributions. In the last part of the section, these definitions and properties are specialized to the optimal filtering context.

**Definition 3.1.** *Two nonnegative measures $\mu, \mu' \in \mathcal{M}^+(E)$ are said* comparable, *if there exist positive constants $0 < a \leq b$, such that*

$$a\, \mu'(A) \leq \mu(A) \leq b\, \mu'(A) ,$$

*for any Borel subset $A \subset E$.*

**Definition 3.2 (Mixing property).** *The nonnegative kernel $K$ defined on $E$ is said* mixing, *if there exist a constant $0 < \varepsilon \leq 1$, and a nonnegative measure $\lambda \in \mathcal{M}^+(E)$, such that*

$$\varepsilon\, \lambda(A) \leq K(x, A) \leq \frac{1}{\varepsilon}\, \lambda(A) ,$$

*for any $x \in E$, and any Borel subset $A \subset E$.*

**Definition 3.3 (Hilbert metric).** *The Hilbert metric on $\mathcal{M}^+(E)$ is defined by*

$$h(\mu, \mu') = \begin{cases} \log \dfrac{\displaystyle\sup_{A\,:\,\mu'(A)>0} \frac{\mu(A)}{\mu'(A)}}{\displaystyle\inf_{A\,:\,\mu'(A)>0} \frac{\mu(A)}{\mu'(A)}} , & \text{if } \mu \text{ and } \mu' \text{ are nonzero and comparable,} \\[2em] +\infty , & \text{otherwise.} \end{cases}$$

Notice that the two nonnegative measures $\mu$ and $\mu'$ are comparable if and only if $\mu$ and $\mu'$ are equivalent, with Radon–Nikodym derivatives $\dfrac{d\mu}{d\mu'}$ and $\dfrac{d\mu'}{d\mu}$ bounded and bounded away from zero, and then the following equality holds

$$h(\mu, \mu') = \log[ \sup_{A\,:\,\mu'(A)>0} \frac{\mu(A)}{\mu'(A)} \sup_{A\,:\,\mu(A)>0} \frac{\mu'(A)}{\mu(A)} ] = \log( \| \frac{d\mu}{d\mu'} \| \, \| \frac{d\mu'}{d\mu} \| ) . \tag{2}$$

Moreover $h$ is a projective distance, i.e. it is invariant under multiplication by positive scalars, hence the Hilbert distance between two unnormalized nonnegative measures is the same as the Hilbert distance between the two corresponding normalized measures : $h(\mu, \mu') = h(\bar{\mu}, \bar{\mu}')$, for any $\mu, \mu' \in \mathcal{M}^+(E)$. In the nonlinear filtering context, this property will allow us to consider the linear transformation $\mu \longmapsto R_n\, \mu$ instead of the nonlinear transformation $\mu \longmapsto \bar{R}_n(\mu) = R_n\, \mu / (R_n\, \mu)(E)$. This projective property does not hold for others distances. Indeed, the following estimates show how the error between two unnormalized nonnegative measures can be used to bound the error between the two corresponding normalized measures. From the decomposition

$$\bar{\mu} - \bar{\mu}' = \frac{1}{\mu(E)}\, [\, \mu - \mu' - (\mu(E) - \mu'(E))\, \bar{\mu}' \,] ,$$

it follows immediately that

$$|\langle \bar{\mu} - \bar{\mu}', \phi \rangle| \leq \frac{|\langle \mu - \mu', \phi \rangle|}{\mu(E)} + \frac{|\mu(E) - \mu'(E)|}{\mu(E)}\, \|\phi\| , \tag{3}$$

and

$$\|\bar{\mu} - \bar{\mu}'\| \leq \frac{\|\mu - \mu'\|}{\mu(E)} + \frac{|\mu(E) - \mu'(E)|}{\mu(E)} , \tag{4}$$

for any $\mu, \mu' \in \mathcal{M}^+(E)$.

The following two lemmas give several useful relations between the Hilbert metric, the total variation norm and a weaker distance suitable for random probability distributions.

**Lemma 3.4.** *For any nonzero $\mu, \mu' \in \mathcal{M}^+(E)$*

$$\|\bar\mu - \bar\mu'\| \leq \frac{2}{\log 3}\, h(\mu, \mu') \; . \tag{5}$$

*If in addition the nonnegative kernel $K$ defined on $E$ is mixing, then*

$$h(K\,\mu, K\,\mu') \leq \frac{1}{\varepsilon^2}\, \|\bar\mu - \bar\mu'\| \; . \tag{6}$$

Proof of Lemma 3.4. The proof of the first inequality can be found in Atar and Zeitouni [3]. To prove the second inequality, notice first that, for any nonzero $\mu, \mu' \in \mathcal{M}^+(E)$

$$\begin{aligned} h(\mu, \mu') &= \log \sup_{A\,:\,\mu'(A)>0} \frac{\mu(A)}{\mu'(A)} + \log \sup_{A\,:\,\mu(A)>0} \frac{\mu'(A)}{\mu(A)} \\ &\leq \sup_{A\,:\,\mu'(A)>0} \frac{|\mu(A) - \mu'(A)|}{\mu'(A)} + \sup_{A\,:\,\mu(A)>0} \frac{|\mu(A) - \mu'(A)|}{\mu(A)} \; , \end{aligned}$$

since $\log(1+x) \leq |x|$. In order to apply this bound to $h(K\,\mu, K\,\mu') = h(K\,\bar\mu, K\,\bar\mu')$, we introduce

$$\begin{aligned} \Delta(A) := \frac{K\,\bar\mu(A) - K\,\bar\mu'(A)}{K\,\bar\mu(A)} &= \int_E (\bar\mu - \bar\mu')(dx)\, \Phi(x, A) \\ &= \int_E (\bar\mu - \bar\mu')^+(dx)\, \Phi(x, A) - \int_E (\bar\mu - \bar\mu')^-(dx)\, \Phi(x, A) \; , \end{aligned}$$

where

$$\Phi(x, A) := \frac{K(x, A)}{K\,\bar\mu(A)} \leq \frac{1}{\varepsilon^2} \; ,$$

for any $x \in E$ and any Borel subset $A \subset E$, using the mixing property. By the Scheffe theorem

$$\int_E (\bar\mu - \bar\mu')^+(dx) = \int_E (\bar\mu - \bar\mu')^-(dx) = \tfrac{1}{2}\, \|\bar\mu - \bar\mu'\| \; ,$$

hence if $\Delta(A)$ is positive, then

$$|\Delta(A)| \leq \int_E (\bar\mu - \bar\mu')^+(dx)\, \Phi(x, A) \leq \frac{1}{2\varepsilon^2}\, \|\bar\mu - \bar\mu'\| \; ,$$

and similarly, if $\Delta(A)$ is negative, then

$$|\Delta(A)| \leq \int_E (\bar\mu - \bar\mu')^-(dx)\, \Phi(x, A) \leq \frac{1}{2\varepsilon^2}\, \|\bar\mu - \bar\mu'\| \; . \quad \Box$$

**Lemma 3.5.** *If the nonnegative kernel $K$ defined on $E$ is* dominated, *i.e. if there exist a constant $c > 0$, and a nonnegative measure $\lambda \in \mathcal{M}^+(E)$, such that*

$$K(x, A) \leq c\, \lambda(A) \; ,$$

*for any $x \in E$, and any Borel subset $A \subset E$, then*

$$\mathbb{E}\, \|K\,\mu - K\,\mu'\| \leq c\, \lambda(E) \sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}\, |\langle \mu - \mu', \phi \rangle| \; ,$$

*for any $\mu, \mu' \in \mathcal{M}^+(E)$, possibly random.*

**Remark 3.6.** If the nonnegative kernel $K$ is mixing, then it is dominated, with the same nonnegative measure $\lambda \in \mathcal{M}^+(E)$, and with $c = 1/\varepsilon$.

**Remark 3.7.** If in addition the nonnegative kernel $K$ is $\mathcal{F}$–measurable, then the same estimate holds for conditional expectations w.r.t. $\mathcal{F}$, i.e.

$$\mathbb{E}[\, \|K\,\mu - K\,\mu'\| \mid \mathcal{F}\,] \leq c\, \lambda(E) \sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\, |\langle \mu - \mu', \phi \rangle| \mid \mathcal{F}\,] \; . \tag{7}$$

Proof of Lemma 3.5. By definition, if $K$ is dominated, then $K(x,\cdot)$ is absolutely continuous w.r.t. $\lambda$, with Radon–Nikodym derivative $k(x,\cdot)$ bounded by $c$, for any $x \in E$. Therefore, the total variation norm $\|K\mu - K\mu'\|$ can be written as an integral as follows

$$\|K\,\mu - K\,\mu'\| = \int_E |\int_E (\mu-\mu')(dx)\,k(x,x')|\,\lambda(dx')\ ,$$

hence, taking expectation yields

$$\begin{aligned}\mathbb{E}\|K\,\mu - K\,\mu'\| &= \int_E \mathbb{E}|\int_E (\mu-\mu')(dx)\,k(x,x')\,|\,\lambda(dx') \\ &\leq \sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}|\langle \mu-\mu',\phi\rangle\,|\,\int_E [\sup_{x\in E} k(x,x')\,]\,\lambda(dx')\ . \quad \square\end{aligned}$$

**Lemma 3.8 (Birkhoff contraction coefficient).** *The nonnegative linear operator on $\mathcal{M}^+(E)$ associated with a nonnegative kernel $K$ defined on $E$, is a contraction under the Hilbert metric, and*

$$\tau(K) := \sup_{0<h(\mu,\mu')<\infty} \frac{h(K\,\mu, K\,\mu')}{h(\mu,\mu')} = \tanh[\tfrac{1}{4}\,H(K)]\ , \tag{8}$$

*where the supremum in*

$$H(K) := \sup_{\mu,\mu'} h(K\,\mu, K\,\mu')\ ,$$

*is over nonzero nonnegative measures : $\tau(K)$ is called the Birkhoff contraction coefficient.*

The proof can be found in Birkhoff [5] or in Hopf [17]. Notice that $H(K) < \infty$ implies $\tau(K) < 1$.

## □ Specialization to the optimal filtering context

The stability results stated in the following sections will in general require that for any $n \geq 1$, the nonnegative kernel $R_n$ is mixing, i.e. there exist a constant $0 < \varepsilon_n \leq 1$, and a nonnegative measure $\lambda_n \in \mathcal{M}^+(E)$, such that

$$\varepsilon_n\,\lambda_n(A) \leq R_n(x,A) \leq \frac{1}{\varepsilon_n}\,\lambda_n(A)\ ,$$

for any $x \in E$, and any Borel subset $A \subset E$. Notice that in full generality $\varepsilon_n$ and $\lambda_n$ depend on the observation $Y_n$, hence are random variables.

**Lemma 3.9.** *The nonnegative linear operator $R_n = \Psi_n\,Q_n$ on $\mathcal{M}^+(E)$ is a contraction under the Hilbert metric, with Birkhoff contraction coefficient $\tau_n := \tau(R_n) \leq 1$. Moreover*

*(i) If $R_n$ is mixing, with the possibly random constant $\varepsilon_n$, then*

$$\tau_n \leq \frac{1-\varepsilon_n^2}{1+\varepsilon_n^2} < 1\ .$$

*(ii) If $Q_n$ is mixing, with the nonrandom constant $\varepsilon_n$, then $R_n$ is also mixing, with the same constant $\varepsilon_n$, and*

$$\tau_n \leq \tau(Q_n) \leq \frac{1-\varepsilon_n^2}{1+\varepsilon_n^2} < 1\ .$$

Throughout the paper, for any integers $m \leq n$, the contraction coefficient of the product $R_{n:m} = R_n \cdots R_m$ is denoted by $\tau_{n:m} := \tau(R_{n:m}) \leq \tau_n \cdots \tau_m$ and by convention $\tau_{n:n+1} = \tau_{m-1:m} = 1$.

Proof of Lemma 3.9. It follows immediately from Lemma 3.8 that $R_n$ is a contraction under the Hilbert metric as a nonnegative linear operator on $\mathcal{M}^+(E)$.

If $R_n$ is mixing, then for any nonzero $\mu,\mu' \in \mathcal{M}^+(E)$, and any Borel subset $A \subset E$

$$\varepsilon_n^2\,\frac{R_n\,\mu'(A)}{\mu'(E)} \leq \varepsilon_n\,\lambda_n(A) \leq \frac{R_n\,\mu(A)}{\mu(E)} \leq \frac{1}{\varepsilon_n}\,\lambda_n(A) \leq \frac{1}{\varepsilon_n^2}\,\frac{R_n\,\mu'(A)}{\mu'(E)}\ ,$$

hence $R_n\,\mu$ and $R_n\,\mu'$ are comparable. Using equation (2) yields

$$H(R_n) = \sup_{\mu,\mu'} h(R_n\,\mu, R_n\,\mu') = \sup_{\mu,\mu'} \log(\|\frac{d(R_n\,\mu)}{d(R_n\,\mu')}\|\,\|\frac{d(R_n\,\mu')}{d(R_n\,\mu)}\|) \leq \log\frac{1}{\varepsilon_n^4}\;,$$

where the supremum is taken over nonzero nonnegative measures. Then using Lemma 3.8 yields

$$\tau_n = \tau(R_n) = \tanh[\tfrac{1}{4}\,H(R_n)] \leq \tanh(\log\frac{1}{\varepsilon_n}) = \frac{1-\varepsilon_n^2}{1+\varepsilon_n^2} < 1\;,$$

which ends the proof of (i).

If $Q_n$ is mixing, then $R_n = \Psi_n\,Q_n$ is also mixing, since

$$\varepsilon_n \int_A \Psi_n(x')\,\lambda_n(dx') \leq R_n(x,A) \leq \frac{1}{\varepsilon_n}\int_A \Psi_n(x')\,\lambda_n(dx')\;,$$

for any $x \in E$, and any Borel subset $A \subset E$, hence for any nonzero $\mu,\mu' \in \mathcal{M}^+(E)$, $R_n\,\mu$ and $R_n\,\mu'$ are comparable, with Radon–Nikodym derivative

$$\frac{d(R_n\,\mu)}{d(R_n\,\mu')}(x') = \frac{d(Q_n\,\mu)}{d(Q_n\,\mu')}(x')\;\mathbf{1}_{\{\Psi_n(x')>0\}} \leq \frac{d(Q_n\,\mu)}{d(Q_n\,\mu')}(x')\;,$$

for any $x' \in E$, and similarly with interchanging the role of $\mu$ and $\mu'$. Therefore

$$H(R_n) \leq \sup_{\mu,\mu'} \log(\|\frac{d(Q_n\,\mu)}{d(Q_n\,\mu')}\|\,\|\frac{d(Q_n\,\mu')}{d(Q_n\,\mu)}\|) = H(Q_n) \leq \log\frac{1}{\varepsilon_n^4}\;,$$

where the supremum is taken over nonzero nonnegative measures. Then using again Lemma 3.8, yields $\tau_n = \tau(R_n) \leq \tau(Q_n)$. □

It follows from (ii) that a sufficient condition for $R_n$ to be mixing, is that $Q_n$ is mixing, but this is not a necessary condition, as illustrated by the example below, where the Markov kernel $Q_n$ is not mixing, but the nonnegative kernel $R_n$ is (equivalent, in a sense to be defined below, to) a mixing kernel.

**Example 3.10.** Assume that $\mu_0$ has compact support $C_0 \subset E$, and that for any $n \geq 1$, the function $\Psi_n$ has compact support $C_n \subset E$, and the transition probability kernel $Q_n$ is defined by

$$Q_n(x,dx') = (2\,\pi)^{-m/2}\;\exp\{-\tfrac{1}{2}\,|x' - f_n(x)|^2\}\,dx' = q_n(x,x')\,\lambda(dx')\;,$$

where the function $f_n$ is continuous, and where

$$\lambda(dx') = (2\,\pi)^{-m/2}\;\exp\{-\tfrac{1}{2}\,|x'|^2\}\,dx'\;.$$

Clearly, the Markov kernel $Q_n$ is not mixing, but introducing

$$\Delta_{n-1} = \sup_{x\in C_{n-1}} |f_n(x)| \qquad\text{and}\qquad \Delta'_n = \sup_{x'\in C_n} |x'|\;,$$

which are both finite a.s., it holds

$$\exp\{-\Delta_{n-1}\,\Delta'_n - \Delta_{n-1}^2\} \leq q_n(x,x') \leq \exp\{\Delta_{n-1}\,\Delta'_n\}\;, \tag{9}$$

for any $x \in C_{n-1}$ and any $x' \in C_n$. Define

$$R_n(x,dx') = Q_n(x,dx')\,\Psi_n(x')\;,$$

as usual, and

$$\begin{aligned} R_n^\bullet(x,dx') &= \mathbf{1}_{\{x\,\in\,C_{n-1}\}}\,R_n(x,dx') + \mathbf{1}_{\{x\,\notin\,C_{n-1}\}}\,\Psi_n(x')\,\lambda(dx') \\ &= [\mathbf{1}_{\{x\,\in\,C_{n-1}\}}\,q_n(x,x') + \mathbf{1}_{\{x\,\notin\,C_{n-1}\}}]\,\Psi_n(x')\,\lambda(dx')\;. \end{aligned}$$

Notice first that the sequence $\{\mu_n \, , \, n \geq 0\}$ defined by (1) satisfies also

$$\mu_n = \frac{R^\bullet_n \, \mu_{n-1}}{(R^\bullet_n \, \mu_{n-1})(E)} \; . \tag{10}$$

Moreover, it follows from (9) that

$$\exp\{-\Delta_{n-1} \, \Delta'_n - \Delta^2_{n-1}\} \int_A \Psi_n(x') \, \lambda(dx') \leq R^\bullet_n(x, A) \leq \exp\{\Delta_{n-1} \, \Delta'_n\} \int_A \Psi_n(x') \, \lambda(dx') \; ,$$

for any $x \in E$, and any Borel subset $A \subset E$, i.e. the nonnegative kernel $R^\bullet_n$ is mixing. Therefore, stability and approximation properties of the sequence $\{\mu_n \, , \, n \geq 0\}$ defined by (1), can be obtained directly by studying (10) instead, which involves mixing operators.

# 4 Stability of nonlinear filters

In practice one has rarely access to the initial distribution of the hidden state process, hence it is important to study the stability of the filter w.r.t. its initial condition. Moreover, the answer to this question will be useful to study the stability of the filter w.r.t. the model.

Let $\mu_n$ denote the filter initialized with the correct $\mu_0$, and let $\mu'_n$ denote the filter initialized with a wrong $\mu'_0$, i.e. $\mu_n = \bar{R}_{n:1}(\mu_0)$ and $\mu'_n = \bar{R}_{n:1}(\mu'_0)$. We are interested in the total variation error at time $n$ induced by the initial error.

**Theorem 4.1.** *Without any assumption on the nonnegative kernels, the following inequality holds*

$$\|\mu_n - \mu'_n\| \leq \frac{2}{\log 3} \, \tau_{n:m} \, h(\mu_{m-1}, \mu'_{m-1}) \; .$$

*If in addition the nonnegative kernel $R_m$ is mixing, then*

$$\|\mu_n - \mu'_n\| \leq \frac{2}{\log 3} \, \tau_{n:m+1} \, \frac{1}{\varepsilon^2_m} \, \|\mu_{m-1} - \mu'_{m-1}\| \; .$$

**Corollary 4.2.** *If for any $k \geq 1$, the nonnegative kernel $R_k$ is mixing with $\varepsilon_k \geq \varepsilon > 0$, then convergence holds uniformly in time, i.e.*

$$\|\mu_n - \mu'_n\| \leq \frac{2}{\varepsilon^2 \, \log 3} \, \tau^{n-m} \, \|\mu_{m-1} - \mu'_{m-1}\| \qquad \textit{with} \qquad \tau := \frac{1-\varepsilon^2}{1+\varepsilon^2} \; .$$

Proof of Theorem 4.1. Using (5), and the definition (8) of the Birkhoff contraction coefficient, yields

$$\|\bar{R}_{n:m}(\mu) - \bar{R}_{n:m}(\mu')\| \leq \frac{2}{\log 3} \, h(R_{n:m} \, \mu, R_{n:m} \, \mu') \leq \frac{2}{\log 3} \, \tau_{n:m} \, h(\mu, \mu') \; , \tag{11}$$

for any $\mu, \mu' \in \mathcal{P}(E)$. If the nonnegative kernel $R_m$ is mixing, then using (6) yields

$$\begin{aligned} \|\bar{R}_{n:m}(\mu) - \bar{R}_{n:m}(\mu')\| &\leq \frac{2}{\log 3} \, h(R_{n:m+1} \, R_m \, \mu, R_{n:m+1} \, R_m \, \mu') \\ &\leq \frac{2}{\log 3} \, \tau_{n:m+1} \, h(R_m \, \mu, R_m \, \mu') \leq \frac{2}{\log 3} \, \tau_{n:m+1} \, \frac{1}{\varepsilon^2_m} \, \|\mu - \mu'\| \; . \end{aligned} \tag{12}$$

Taking $\mu = \mu_m$ and $\mu' = \mu'_m$ finishes the proof. □

To solve the nonlinear filtering problem, one must have a model to describe the state / observation system, $\{X_n \, , \, n \geq 0\}$, $\{Y_n \, , \, n \geq 1\}$, as presented in Section 2. The general hidden Markov model is based on the initial condition $\mu_0$, on the transition kernels $Q_n$ and on the likelihood functions $\Psi_n$, which define the evolution operator $R_n$ for the optimal filter $\mu_n$. But, as for the initial condition, in practice one has rarely access to the true model. In particular, the prior information on the state sequence is in general unknown and the choice of $Q_n$ is approximative. Similarly, the probabilistic relation between the observation and the state is in general unknown and the choice of $\Psi_n$ is also approximative. As a result, instead of using the true model, it is common to work with a wrong model, based on a wrong transition kernel $Q'_n$ and a wrong likelihood function $\Psi'_n$, which define the evolution operator $R'_n$ for a wrong filter $\mu'_n$.

Another situation is when the evolution operator $R_n$ is known, but difficult to compute. For the purpose of practical implementation, one constructs an approximate filter $\mu'_n$ such that the evolution $\mu'_{n-1} \longmapsto \mu'_n$ is easy to compute and close to the true evolution $\mu'_{n-1} \longmapsto \bar{R}_n(\mu'_{n-1})$.

We are interested in bounding the *global error* between $\mu'_n$ and $\mu_n$ induced by the *local errors* committed at each time step. We suppose here that $\mu_0 = \mu'_0$, since the problem of a wrong initialization has already been studied above. In full generality, we assume that $\{\mu'_n\,,\, n \geq 0\}$ is a random sequence with values in $\mathcal{P}(E)$, satisfying the following property : for any $n \geq k \geq 1$ and for any bounded measurable function $F$ defined on $\mathcal{P}(E)$

$$\mathbb{E}[F(\mu'_k) \mid Y_{1:n}] = \mathbb{E}[F(\mu'_k) \mid Y_{1:k}] \,. \tag{13}$$

The results stated below are based on the following decomposition of the global error into sums of local errors transported by a sequence of normalized evolution operators $\bar{R}_n$,

$$\mu'_n - \mu_n = \sum_{k=1}^{n} [\, \bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k}(\mu'_{k-1})\,] = \sum_{k=1}^{n} [\, \bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})\,] \,. \tag{14}$$

This equation shows the close relation between the stability w.r.t. the initial condition and the stability w.r.t. the model.

Let us consider first the case where we can estimate the local error in the sense of the Hilbert metric.

**Assumption H** (local error bound in the Hilbert metric) **:**

$$\delta_k^{\mathrm{H}} := \mathbb{E}[\, h(\mu'_k\,, \bar{R}_k(\mu'_{k-1})) \mid Y_{1:k}\,] < \infty \ .$$

**Remark 4.3.** If the evolution of the wrong filter $\mu'_k$ is defined by another deterministic nonnegative kernel $R'_k(x,dx') = Q'_k(x,dx')\, \Psi'_k(x')$, and if

$$Q_k(x,dx') = q_k(x,x')\, \lambda_k(dx') \qquad \text{and} \qquad Q'_k(x,dx') = q'_k(x,x')\, \lambda_k(dx') \ ,$$

then a sufficient condition for Assumption H to hold is that there exist $\delta_k \geq 0$ and $a_k > 0$, such that

$$a_k \leq \frac{\Psi_k(x')\, q_k(x,x')}{\Psi'_k(x')\, q'_k(x,x')} \leq a_k\, \exp(\delta_k) \ ,$$

for all $x, x' \in E$, in which case $\delta_k^{\mathrm{H}} \leq \delta_k$.

**Theorem 4.4.** *If for any $k \geq 1$, Assumption H holds, then*

$$\mathbb{E}[\, \|\mu'_n - \mu_n\| \mid Y_{1:n}\,] \leq \frac{2}{\log 3} \sum_{k=1}^{n} \tau_{n:k+1}\, \delta_k^{\mathrm{H}} \ . \tag{15}$$

**Corollary 4.5.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing with $\varepsilon_k \geq \varepsilon > 0$, and Assumption H holds with $\delta_k^{\mathrm{H}} \leq \delta$, then convergence holds uniformly in time, i.e.*

$$\mathbb{E}[\, \|\mu_n - \mu'_n\| \mid Y_{1:n}\,] \leq \frac{2}{\varepsilon^2\, \log 3}\, \delta \ . \tag{16}$$

Indeed, (16) follows from

$$\sum_{k=1}^{n} \tau^{n-k} = \frac{1-\tau^n}{1-\tau} \leq \frac{1}{1-\tau} = \frac{1+\varepsilon^2}{2\,\varepsilon^2} \leq \frac{1}{\varepsilon^2} \ .$$

Proof of Theorem 4.4. Using the decomposition (14), the triangle inequality, and estimate (11), yields

$$\|\mu'_n - \mu_n\| \leq \sum_{k=1}^{n} \|\bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})\| \leq \frac{2}{\log 3} \sum_{k=1}^{n} \tau_{n:k+1}\, h(\mu'_k, \bar{R}_k(\mu'_{k-1})) \ .$$

Taking conditional expectation w.r.t. the observations and using (13), yields (15). □

Let us consider next the case where we can estimate the local error in the sense of the total variation norm

$$\delta_k^{\mathrm{TV}} := \mathbb{E}[\, \|\mu'_k - \bar{R}_k(\mu'_{k-1})\| \mid Y_{1:k}\,] \leq 2 \ .$$

**Theorem 4.6.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing, then*

$$\mathbb{E}[\, \|\mu_n - \mu'_n\| \mid Y_{1:n}\,] \leq \delta_n^{\mathrm{TV}} + \frac{2}{\log 3} \sum_{k=1}^{n-1} \tau_{n:k+2}\, \frac{\delta_k^{\mathrm{TV}}}{\varepsilon_{k+1}^2} \; . \tag{17}$$

**Corollary 4.7.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing with $\varepsilon_k \geq \varepsilon > 0$, and $\delta_k^{\mathrm{TV}} \leq \delta$, then convergence holds uniformly in time, i.e.*

$$\mathbb{E}[\, \|\mu_n - \mu'_n\| \mid Y_{1:n}\,] \leq (\, 1 + \frac{2}{\varepsilon^4 \log 3}\,)\, \delta \; .$$

Proof of Theorem 4.6. The decomposition (14) is written as

$$\mu'_n - \mu_n = [\, \mu'_n - \bar{R}_n(\mu'_{n-1})\,] + \sum_{k=1}^{n-1} [\, \bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})\,] \; , \tag{18}$$

hence using the triangle inequality and estimate (12), yields

$$\begin{aligned} \|\mu'_n - \mu_n\| \;&\leq\; \|\mu'_n - \bar{R}_n(\mu'_{n-1})\| + \sum_{k=1}^{n-1} \|\bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})\| \\ &\leq\; \|\mu'_n - \bar{R}_n(\mu'_{n-1})\| + \frac{2}{\log 3} \sum_{k=1}^{n-1} \tau_{n:k+2}\, \frac{1}{\varepsilon_{k+1}^2}\, \|\mu'_k - \bar{R}_k(\mu'_{k-1})\| \; . \end{aligned}$$

Taking conditional expectation w.r.t. the observations and using (13), yields(17). □

Let us consider finally the case where we can only estimate the local error in the weak sense

$$\delta_k^{\mathrm{W}} := \sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, |\, \langle \mu'_k - \bar{R}_k(\mu'_{k-1}), \phi \rangle \,| \mid Y_{1:k}\,] \leq 2 \; .$$

This typically happens if the approximate filter $\mu'_k$ is an empirical probability distribution associated with $\bar{R}_k(\mu'_{k-1})$ : in this case, bounding the local error requires to use the law of large numbers, which can only provide estimates in the weak sense. However, if the nonnegative kernel $R_{k+1}$ is dominated, then using Lemma 3.5, the local error transported by $R_{k+1}$ can be bounded in total variation with the same precision $\delta_k^{\mathrm{W}}$ as in the weak sense.

**Theorem 4.8.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing, then*

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, |\, \langle \mu_n - \mu'_n, \phi \rangle \,| \mid Y_{1:n}\,] \leq \delta_n^{\mathrm{W}} + 2\, \frac{\delta_{n-1}^{\mathrm{W}}}{\varepsilon_n^2} + \frac{4}{\log 3} \sum_{k=1}^{n-2} \tau_{n:k+3}\, \frac{\delta_k^{\mathrm{W}}}{\varepsilon_{k+2}^2\, \varepsilon_{k+1}^2} \; . \tag{19}$$

**Corollary 4.9.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing with $\varepsilon_k \geq \varepsilon > 0$, and $\delta_k^{\mathrm{W}} \leq \delta$, then convergence holds uniformly in time, i.e.*

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, |\, \langle \mu_n - \mu'_n, \phi \rangle \,| \mid Y_{1:n}\,] \leq (\, 1 + \frac{2}{\varepsilon^2} + \frac{4}{\varepsilon^6 \log 3}\,)\, \delta \; .$$

Proof of Theorem 4.8. Using the decomposition (18) and the triangle inequality, yields

$$\begin{aligned} |\, \langle \mu'_n - \mu_n, \phi \rangle \,| \;&\leq\; |\, \langle \mu'_n - \bar{R}_n(\mu'_{n-1}), \phi \rangle \,| \\ &\quad + \sum_{k=1}^{n-1} \|\, \bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})\,\|\; \|\phi\| \; . \end{aligned} \tag{20}$$

For any $1 \leq k \leq n-2$, using estimate (12) yields

$$\begin{aligned} \|\bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})\,\| \;&=\; \|\bar{R}_{n:k+2} \circ \bar{R}_{k+1}(\mu'_k) - \bar{R}_{n:k+2} \circ \bar{R}_{k+1} \circ \bar{R}_k(\mu'_{k-1})\,\| \\ &\leq\; \frac{2}{\log 3}\, \tau_{n:k+3}\, \frac{1}{\varepsilon_{k+2}^2}\, \|\bar{R}_{k+1}(\mu'_k) - \bar{R}_{k+1} \circ \bar{R}_k(\mu'_{k-1})\,\| \; . \end{aligned}$$

For any $1 \leq k \leq n-1$, using estimate (4) yields

$$\| \bar{R}_{k+1}(\mu'_k) - \bar{R}_{k+1} \circ \bar{R}_k(\mu'_{k-1}) \| \leq 2 \, \frac{\| R_{k+1} \, (\mu'_k - \bar{R}_k(\mu'_{k-1})) \|}{(R_{k+1} \, \mu'_k)(E)} \; ,$$

and the mixing property yields

$$(R_{k+1} \, \mu'_k)(E) \geq \varepsilon_{k+1} \, \lambda_{k+1}(E) \; .$$

Taking conditional expectation w.r.t. the observations, using estimate (7) with $K = R_{k+1}$, $\mu = \bar{R}_k(\mu'_{k-1})$, $\mu' = \mu'_k$ and $\mathcal{F} = Y_{1:n}$, and using (13), yields

$$\begin{aligned} \mathbb{E}[\, \| R_{k+1} \, (\mu'_k - \bar{R}_k(\mu'_{k-1})) \| \mid Y_{1:n} \,] \; &\leq \; \frac{\lambda_{k+1}(E)}{\varepsilon_{k+1}} \sup_{\phi \, : \, \|\phi\|=1} \mathbb{E}[\, | \langle \mu'_k - \bar{R}_k(\mu'_{k-1}), \phi \rangle | \mid Y_{1:n} \,] \\ &\leq \; \frac{\lambda_{k+1}(E)}{\varepsilon_{k+1}} \, \delta_k^{\mathrm{W}} \; . \end{aligned}$$

Combining these estimates yields

$$\mathbb{E}[\, \| \bar{R}_{k+1}(\mu'_k) - \bar{R}_{k+1} \circ \bar{R}_k(\mu'_{k-1}) \| \mid Y_{1:n} \,] \leq 2 \, \frac{\delta_k^{\mathrm{W}}}{\varepsilon_{k+1}^2} \; .$$

Finally, taking conditional expectation w.r.t. the observations in (20), yields (19). □

# 5 Uniform convergence of interacting particle filters

In this section and in the next section, we consider again the framework introduced in Section 4, but now the wrong model is chosen deliberately, such that the wrong filter can easily be computed, and remains close to the optimal filter. More specifically, we are interested in particle methods to approximate numerically the optimal filter, and we provide estimates of the approximation error. The idea common to all particle filters is to generate an $N$–sample $(\xi^1_{n|n-1}, \cdots, \xi^N_{n|n-1})$ of i.i.d. random variables, called a *particle system*, with common probability distribution $Q_n \, \mu^N_{n-1}$, where $\mu^N_{n-1}$ is an approximation of $\mu_{n-1}$, and to use the corresponding empirical probability distribution

$$\mu^N_{n|n-1} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i_{n|n-1}} \; ,$$

as an approximation of $\mu_{n|n-1} = Q_n \, \mu_{n-1}$. The method is very easy to implement, even in high dimensional problems, since it is sufficient in principle to simulate independent samples of the hidden state sequence. A major and earliest contribution in this field was made by Gordon, Salmond and Smith [15], which proposed to use sampling / importance resampling (SIR) techniques in the correction step : the positive effect of the resampling step is to automatically select particles with larger values of the likelihood function, i.e. to concentrate particles in regions of interest of the state space. A very complete account of the currently available mathematical results can be found in the survey paper by Del Moral and Miclo [11]. Theoretical and practical aspects can be found in the volume edited by Doucet, de Freitas and Gordon [14].

Throughout the paper, $S^N(\mu)$ is a shorthand notation for the empirical probability distribution of an $N$–sample with probability distribution $\mu$, i.e.

$$S^N(\mu) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i} \qquad \text{with} \qquad (\xi^1, \cdots, \xi^N) \text{ i.i.d.} \sim \mu.$$

**Lemma 5.1.** *For any $\mu \in \mathcal{P}(E)$*

$$\sup_{\phi \, : \, \|\phi\|=1} \mathbb{E}| \, \langle S^N(\mu) - \mu, \phi \rangle \, | \leq \frac{1}{\sqrt{N}} \; .$$

Proof. It holds

$$\langle S^N(\mu) - \mu, \phi \rangle = \frac{1}{N} \sum_{i=1}^N [\phi(\xi^i) - \langle \mu, \phi \rangle] \; ,$$

hence

$$\mathbb{E}| \, \langle S^N(\mu) - \mu, \phi \rangle \, |^2 = \frac{1}{N} \, [\, \langle \mu, \phi^2 \rangle - \langle \mu, \phi \rangle^2 \,] \leq \frac{1}{N} \, \|\phi\|^2 \; . \quad \square$$

**Remark 5.2.** If in addition $\phi$ and $\mu$ are $\mathcal{F}$–measurable r.v.'s, and if conditionally w.r.t. $\mathcal{F}$ the r.v.'s $(\xi^1, \cdots, \xi^N)$ are i.i.d. with (conditional) probability distribution $\mu$, then the same estimate holds for conditional expectation w.r.t. $\mathcal{F}$, i.e.

$$\mathbb{E}[\, | \, \langle S^N(\mu) - \mu, \phi \rangle \, | \; | \; \mathcal{F}] \leq \frac{1}{\sqrt{N}} \, \|\phi\| \; . \tag{21}$$

To approximate the posterior probability distribution $\Lambda \cdot \mu$, it follows immediately from Lemma 5.1, and using estimate (3), that

$$\sup_{\phi \, : \, \|\phi\|=1} \mathbb{E}|\langle \Lambda \cdot S^N(\mu) - \Lambda \cdot \mu, \phi \rangle| \leq 2 \sup_{\phi \, : \, \|\phi\|=1} \mathbb{E}|\frac{\langle S^N(\mu) - \mu, \Lambda \, \phi \rangle}{\langle \mu, \Lambda \rangle}| \leq \frac{2}{\sqrt{N}} \, \frac{\sup_{x \in E} \Lambda(x)}{\langle \mu, \Lambda \rangle} \; .$$

However, it is usually difficult to have a reliable lower bound for the denominator $\langle \mu, \Lambda \rangle$, and the following procedure, classical in sequential analysis, can be used instead.

**Lemma 5.3.** *Let $\mu \in \mathcal{P}(E)$, and let $\Lambda$ be a nonnegative bounded measurable function defined on $E$, such that $\langle \mu, \Lambda \rangle > 0$. For any $\delta > 0$, define the stopping time*

$$T = \inf\{N \, : \, \delta^2 \sum_{i=1}^N \Lambda(\xi^i) \geq \sup_{x \in E} \Lambda(x)\} \qquad \textit{with} \qquad (\xi^1, \cdots, \xi^N, \cdots) \ \textit{i.i.d.} \sim \mu .$$

*Then*

$$\sup_{\phi \, : \, \|\phi\|=1} \mathbb{E}|\langle \Lambda \cdot S^T(\mu) - \Lambda \cdot \mu, \phi \rangle| \leq 2 \, \delta \, \sqrt{1 + \delta^2} \; .$$

*Moreover, the expected number of particles required to obtain an error of order $O(\delta)$, is of order $O(1/\delta^2)$, i.e.*

$$\frac{\rho}{\delta^2} \leq \mathbb{E}[T] \leq \frac{\rho}{\delta^2} \, (1 + \delta^2) \; ,$$

*where* $\rho = \dfrac{\sup_{x \in E} \Lambda(x)}{\langle \mu, \Lambda \rangle}$.

The method proposed here to approximate the posterior probability distribution $\Lambda \cdot \mu$ is somehow intermediate, between the classical importance sampling method, which uses a fixed number of random variables, and the acceptance / rejection method, which requires a random number of random variables. In Lemma 5.3, the number of random variables generated is random as in the acceptance / rejection method, but there is no rejection, since all the random variables generated are explicitly used in the approximation, as in the importance sampling method.

Proof of Lemma 5.3. Notice first that a.s.

$$\frac{1}{N} \sum_{i=1}^N \Lambda(\xi^i) \longrightarrow \langle \mu, \Lambda \rangle > 0 \; ,$$

as $N \uparrow \infty$, hence the stopping time $T$ is a.s. finite. Using estimate (3), yields

$$|\langle \Lambda \cdot S^N(\mu) - \Lambda \cdot \mu, \phi \rangle| \leq 2 \, |\frac{\langle S^N(\mu) - \mu, \Lambda \, \phi \rangle}{\langle S^N(\mu), \Lambda \rangle}| \; ,$$

and we define the ratio

$$R_N = \frac{\langle S^N(\mu) - \mu, \Lambda\, \phi \rangle}{\langle S^N(\mu), \Lambda \rangle} = \frac{M_N}{D_N} \;,$$

where

$$M_N = \sum_{i=1}^N [\, \Lambda(\xi^i)\, \phi(\xi^i) - \langle \mu, \Lambda\, \phi \rangle \,] \qquad \text{and} \qquad D_N = \sum_{i=1}^N \Lambda(\xi^i) \;.$$

The sequence $\{M_N \,,\, N \geq 1\}$ is a $\{\mathcal{F}_N \,,\, N \geq 1\}$–martingale, where $\mathcal{F}_N = \sigma(\xi^1, \cdots, \xi^N)$, and let $M_N^2 = V_N + M'_N$ be the Doob decomposition of the submartingale $\{M_N^2 \,,\, N \geq 1\}$, where the sequence $\{M'_N \,,\, N \geq 1\}$ is a martingale, and

$$V_N = \sum_{i=1}^N \mathbb{E}[\, (\Delta M_i)^2 \mid \mathcal{F}_{i-1} \,] = N\, [\, \langle \mu, \Lambda^2\, \phi^2 \,\rangle - \langle \mu, \Lambda\, \phi \rangle^2 \,] \leq N\, \|\phi\|^2\, \lambda\, \langle \mu, \Lambda \rangle \;,$$

where $\lambda = \sup_{x \in E} \Lambda(x)$. By definition of the stopping time $T$, it holds

$$\frac{\lambda}{\delta^2} \leq D_T = D_{T-1} + \Lambda(\xi^T) \;\leq \frac{\lambda}{\delta^2} + \lambda = \frac{\lambda}{\delta^2}\, (1 + \delta^2) \;,$$

which yields

$$|R_T| \leq \frac{\delta^2}{\lambda}\, |M_T| \;.$$

The Cauchy–Schwartz inequality yields

$$\mathbb{E}|R_T| \leq \frac{\delta^2}{\lambda}\, \mathbb{E}|M_T| \leq \frac{\delta^2}{\lambda}\, (\mathbb{E}[M_T^2])^{1/2} = \frac{\delta^2}{\lambda}\, (\mathbb{E}[V_T] + \mathbb{E}[M'_T])^{1/2} \;.$$

The optional stopping theorem yields

$$\mathbb{E}[M'_T] = 0 \qquad \text{and} \qquad \mathbb{E}[V_T] \leq \|\phi\|^2\, \lambda\, \langle \mu, \Lambda \rangle\, \mathbb{E}[T] \;.$$

The Wald identity yields

$$\langle \mu, \Lambda \rangle\, \mathbb{E}[T] = \mathbb{E}[D_T] \leq \frac{\lambda}{\delta^2}\, (1 + \delta^2) \;,$$

hence

$$\mathbb{E}|R_T| \leq \|\phi\|\, \delta\, \sqrt{1 + \delta^2} \qquad \text{and} \qquad \frac{\rho}{\delta^2} \leq \mathbb{E}[T] \leq \frac{\rho}{\delta^2}\, \sqrt{1 + \delta^2} \;. \quad \Box$$

**Remark 5.4.** If in addition $\phi$ and $\mu$ are $\mathcal{F}$–measurable r.v.'s, and if conditionally w.r.t. $\mathcal{F}$ the r.v.'s $(\xi^1, \cdots, \xi^N, \cdots)$ are i.i.d. with (conditional) probability distribution $\mu$, then the same estimate holds for conditional expectation w.r.t. $\mathcal{F}$, i.e.

$$\mathbb{E}[\, |\langle \Lambda \cdot S^T(\mu) - \Lambda \cdot \mu, \phi \rangle| \mid \mathcal{F}] \leq 2\, \delta\, \sqrt{1 + \delta^2}\; \|\phi\| \;, \tag{22}$$

and

$$\frac{\rho}{\delta^2} \leq \mathbb{E}[T \mid \mathcal{F}] \leq \frac{\rho}{\delta^2}\, (1 + \delta^2) \;.$$

## □ Interacting particle filter

Let $\mu_n^N$ denote the interacting particle filter (IPF) approximation of $\mu_n$. The transition from $\mu_{n-1}^N$ to $\mu_n^N$ is described by the following diagram

$$\mu_{n-1}^N \xrightarrow[\text{sampled prediction}]{} \mu_{n|n-1}^N = S^N(Q_n\, \mu_{n-1}^N) \xrightarrow[\text{correction}]{} \mu_n^N = \Psi_n \cdot \mu_{n|n-1}^N \;.$$

In practice, the particle approximation $\mu_{n|n-1}^N$ is completely characterized by the particle system $(\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^N)$, and the transition from $(\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^N)$ to $(\xi_{n+1|n}^1, \cdots, \xi_{n+1|n}^N)$ consists of the following three steps.

(i) Correction : for all $i = 1, \cdots, N$, compute the weight

$$\omega_n^i = \frac{1}{c_n} \Psi_n(\xi_{n|n-1}^i) ,$$

with the normalization constant

$$c_n = \sum_{i=1}^{N} \Psi_n(\xi_{n|n-1}^i) .$$

Then set

$$\mu_n^N = \Psi_n \cdot \mu_{n|n-1}^N = \sum_{i=1}^{N} \omega_n^i \, \delta_{\xi_{n|n-1}^i} .$$

(ii) Resampling : independently for all $i = 1, \cdots, N$, generate a r.v. $\xi_n^i \sim \mu_n^N$.

(iii) Prediction : independently for all $i = 1, \cdots, N$, generate a r.v. $\xi_{n+1|n}^i \sim Q_{n+1}(\xi_n^i, \cdot)$. Then set

$$\mu_{n+1|n}^N = S^N(Q_{n+1}\, \mu_n^N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_{n+1|n}^i} .$$

The resampling step (ii) requires to generate random variables according to the weighted discrete probability distribution $\mu_n^N$, which can be easily implemented.

First we check that the IPF satisfies (13). Indeed, for any $n \geq k \geq 1$ and for any bounded measurable function $F$ defined on $\mathcal{P}(E)$

$$\mathbb{E}[F(\mu_k^N) \mid Y_{1:n}] = \mathbb{E}[F(\mu_k^N) \mid Y_{1:k}] .$$

**Remark 5.5.** Notice that the weights $\{\omega_n^1, \cdots, \omega_n^N\}$ are not well defined in the particular case where the normalization constant $c_n$ is zero. Hence Del Moral and Jacod [10] propose to reinitialize the particles to a single arbitrary point of the state space whenever $c_n$ is zero, and they study the behavior of the resulting particle system. Obvously, this problem cannot happen if the likelihood function $\Psi_n$ is positive. By construction, the sequential particle filter defined at the end of this section does not run into this problem.

**Remark 5.6.** If the nonnegative operator $R_n$ is mixing, then

$$\inf_{\mu \in \mathcal{P}(E)} \langle Q_n\, \mu, \Psi_n \rangle = \inf_{\mu \in \mathcal{P}(E)} (R_n\, \mu)(E) \geq \varepsilon_n^2\, (R_n\, \mu_{n-1})(E) = \varepsilon_n^2\, \langle \mu_{n|n-1}, \Psi_n \rangle,$$

hence a.s.

$$\inf_{\mu \in \mathcal{P}(E)} \langle Q_n\, \mu, \Psi_n \rangle > 0 ,$$

in view of Remark 2.1.

Without loss of generality, it is assumed that the likelihood function is bounded.

**Assumption L :**

$$\sup_{x \in E} \Psi_k(x) < \infty .$$

If Assumption L holds, and if for any $k \geq 1$, the nonnegative operator $R_k$ is mixing, then the following notation is introduced

$$\rho_k := \frac{\displaystyle\sup_{x \in E} \Psi_k(x)}{\displaystyle\inf_{\mu \in \mathcal{P}(E)} \langle Q_k\, \mu, \Psi_k \rangle} ,$$

and in view of Remark 5.6, $\rho_k$ is a.s. finite.

**Theorem 5.7.** *If for any $k \geq 1$, Assumption L holds, and the nonnegative operator $R_k$ is mixing, then the IPF estimator satisfies*

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n - \mu_n^N\,,\phi\rangle\,|\mid Y_{1:n}\,] \leq \delta_n^{\rm W} + 2\,\frac{\delta_{n-1}^{\rm W}}{\varepsilon_n^2} + \frac{4}{\log 3}\sum_{k=1}^{n-2} \tau_{n:k+3}\,\frac{\delta_k^{\rm W}}{\varepsilon_{k+2}^2\,\varepsilon_{k+1}^2}\ ,$$

*where for any $k \geq 1$*

$$\delta_k^{\rm W} \leq \frac{1}{\sqrt{N}}\,2\,\rho_k\ .$$

**Remark 5.8.** If the transition kernel $Q_{n+1}$ is dominated, i.e. $Q_{n+1}(x,\cdot)$ is absolutely continuous w.r.t. $\lambda_{n+1} \in \mathcal{M}^+(E)$, with density $q_{n+1}(x,\cdot)$ bounded by $c_{n+1}$, for any $x \in E$, then convergence in the weak sense of the particle filter can be used to prove convergence in total variation of the particle predictor. Indeed, using Lemma 3.5 yields

$$\mathbb{E}[\,\|\mu_{n+1|n} - Q_{n+1}\,\mu_n^N\|\mid Y_{1:n}\,] \leq c_{n+1}\,\lambda_{n+1}(E)\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n - \mu_n^N\,,\phi\rangle\,|\mid Y_{1:n}\,]\ ,$$

where both $\mu_{n+1|n}$ and $Q_{n+1}\,\mu_n^N$ are absolutely continuous w.r.t. $\lambda_{n+1}$, and

$$\frac{d(Q_{n+1}\,\mu_n^N)}{d\lambda_{n+1}}(x') = \sum_{i=1}^N \omega_i^N\,q_{n+1}(\xi_{n|n-1}^i, x')\ ,$$

for any $x' \in E$, which can be easily computed.

**Remark 5.9.** In general, it is not realistic to assume that the r.v. $\rho_k$ is a.s. bounded, hence it seems difficult to guarantee that convergence holds uniformly in time, for a given observation sequence. On the other hand, averaging over observation sequences makes it possible to obtain convergence uniformly in time, under more realistic assumptions. Indeed, if for any $k \geq 1$, the nonnegative operator $R_k$ is mixing with nonrandom $\varepsilon_k$, and $\mathbb{E}[\rho_k]$ is finite, then

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}|\,\langle \mu_n - \mu_n^N\,,\phi\rangle\,| \leq \delta_n + 2\,\frac{\delta_{n-1}}{\varepsilon_n^2} + \frac{4}{\log 3}\sum_{k=1}^{n-2} \tau_{n:k+3}\,\frac{\delta_k}{\varepsilon_{k+2}^2\,\varepsilon_{k+1}^2}\ ,$$

where for any $k \geq 1$

$$\delta_k = \mathbb{E}[\delta_k^{\rm W}\,] \leq \frac{1}{\sqrt{N}}\,2\,\mathbb{E}[\rho_k]\ .$$

**Remark 5.10.** Notice that, if the nonnegative operator $R_k$ is mixing, then

$$\frac{\sup_{x\in E} \Psi_k(x)}{\langle \mu_{k|k-1}\,,\Psi_k\rangle} \leq \rho_k \leq \frac{\sup_{x\in E} \Psi_k(x)}{\varepsilon_k^2\,\langle \mu_{k|k-1}\,,\Psi_k\rangle}\ ,$$

and it follows from Remark 2.2 that

$$\mathbb{E}[\,\frac{\sup_{x\in E} \Psi_k(x)}{\langle \mu_{k|k-1}\,,\Psi_k\rangle}\mid Y_{1:k-1}] = \int_F [\,\sup_{x\in E} g_k(x,y)\,]\,\lambda_k^F(dy)\ ,$$

hence a necessary and sufficient condition for $\mathbb{E}[\rho_k]$ to be finite, is

$$\int_F [\,\sup_{x\in E} g_k(x,y)\,]\,\lambda_k^F(dy) < \infty\ .$$

**Corollary 5.11.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing with $\varepsilon_k \geq \varepsilon > 0$ and nonrandom $\varepsilon$, and $\mathbb{E}[\rho_k] \leq \rho$, then convergence, averaged over observation sequences, holds uniformly in time, i.e.*

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}|\,\langle \mu_n - \mu_n^N\,,\phi\rangle\,| \leq (\,1 + \frac{2}{\varepsilon^2} + \frac{4}{\varepsilon^6\,\log 3}\,)\,\delta\ ,$$

*with*

$$\delta \leq \frac{1}{\sqrt{N}}\,2\,\rho\ .$$

Proof of Theorem 5.7. It is sufficient to bound the local error $\delta_k^{\rm W}$ in the weak sense, and to apply Theorem 4.8. Using estimate (3) yields

$$\begin{aligned} |\langle \mu_k^N - \bar{R}_k(\mu_{k-1}^N), \phi\rangle | &= |\langle \Psi_k \cdot (S^N(Q_k\,\mu_{k-1}^N)) - \Psi_k \cdot (Q_k\,\mu_{k-1}^N), \phi\rangle | \\ &\leq \frac{|\langle S^N(Q_k\,\mu_{k-1}^N) - Q_k\,\mu_{k-1}^N, \Psi_k\,\phi\rangle |}{\langle Q_k\,\mu_{k-1}^N, \Psi_k\rangle} \\ &\quad + \frac{|\langle S^N(Q_k\,\mu_{k-1}^N) - Q_k\,\mu_{k-1}^N, \Psi_k\rangle |}{\langle Q_k\,\mu_{k-1}^N, \Psi_k\rangle}\, \|\phi\| \;, \end{aligned} \tag{23}$$

for any bounded measurable test function $\phi$ defined on $E$. By definition

$$\langle Q_k\,\mu_{k-1}^N, \Psi_k\rangle \geq \inf_{\mu\in\mathcal{P}(E)} \langle Q_k\,\mu, \Psi_k\rangle \;.$$

Using estimate (21) with $\Psi_k\,\phi$ instead of $\phi$, $\mu = Q_k\,\mu_{k-1}^N$ and $\mathcal{F} = \sigma(Y_{1:k}\;,\;\mu_{k-1}^N)$, yields

$$\mathbb{E}[\,|\langle S^N(Q_k\,\mu_{k-1}^N) - Q_k\,\mu_{k-1}^N, \Psi_k\,\phi\rangle |\mid Y_{1:k}\;,\;\mu_{k-1}^N] \leq \frac{1}{\sqrt{N}}\,\|\phi\|\,\sup_{x\in E}\Psi_k(x) \;. \qquad \square$$

In the proof of Theorem 5.7, if we use $\langle S^N(Q_k\,\mu_{k-1}^N), \Psi_k\rangle$ instead of $\langle Q_k\,\mu_{k-1}^N, \Psi_k\rangle$ as the denominator in equation (23), we see that, for the local error to be small, the empirical mean of the likelihood function over the predicted particle system should be large enough. This theoretical argument is also supported by numerical evidence, in cases where the likelihood function is localized in a small region of the state space (which typically arises when measurements are accurate). Indeed, such a region can be so small that it does not contain enough points of the predicted particle system, which automatically results in a small value of the predicted empirical mean of the likelihood function. This phenomenon is called *degeneracy of particle weights* and is a known cause of divergence of particle filters. To solve this degeneracy problem, one idea is to add a regularization step to the algorithm : the resulting filters, called regularized particle filters (RPF) are studied in the next section. Another idea is to control the predicted empirical mean

$$\langle S^N(Q_k\,\mu_{k-1}^N), \Psi_k\rangle = \frac{1}{N}\sum_{i=1}^N \Psi_k(\xi_{k|k-1}^i) \;,$$

by using an adaptive number of particles. To guarantee a local error of order $\delta_k$, independently of any lower bound assumption on the likelihood function, we choose a random number of particles

$$N_k := \inf\{N \,:\, \delta_k^2 \sum_{i=1}^N \Psi_k(\xi_{k|k-1}^i) \geq \sup_{x\in E}\Psi_k(x)\} \;, \tag{24}$$

that will automatically fit the difficult case of localized likelihood functions : the resulting filter, called sequential particle filter (SPF) is studied below.

## □ Sequential particle filter

Let $\mu_n^{N_n}$ denote the sequential particle filter (SPF) approximation of $\mu_n$. The transition from $\mu_{n-1}^{N_{n-1}}$ to $\mu_n^{N_n}$ is described by the following diagram

$$\mu_{n-1}^{N_{n-1}} \xrightarrow[\text{sequential sampled prediction}]{} \mu_{n|n-1}^{N_n} = S^{N_n}(Q_n\,\mu_{n-1}^{N_{n-1}}) \xrightarrow[\text{correction}]{} \mu_n^{N_n} = \Psi_n \cdot \mu_{n|n-1}^{N_n} \;.$$

In practice, the particle approximation $\mu_{n|n-1}^{N_n}$ is completely characterized by the particle system $(\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^{N_n})$, and the transition from $(\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^{N_n})$ to $(\xi_{n+1|n}^1, \cdots, \xi_{n+1|n}^{N_{n+1}})$ consists of the following four steps.

(i) Correction : for all $i = 1, \cdots, N_n$, compute the weight

$$\omega_n^i = \frac{1}{c_n}\, \Psi_n(\xi_{n|n-1}^i) \ ,$$

with the normalization constant

$$c_n = \sum_{i=1}^{N_n} \Psi_n(\xi_{n|n-1}^i) \ .$$

Then set

$$\mu_n^{N_n} = \Psi_n \cdot \mu_{n|n-1}^{N_n} = \sum_{i=1}^{N_n} \omega_n^i \, \delta_{\xi_{n|n-1}^i} \ .$$

(ii) Resampling : independently for all $i = 1, \cdots, N, \cdots$, generate a r.v. $\xi_n^i \sim \mu_n^{N_n}$.

(iii) Prediction : independently for all $i = 1, \cdots, N, \cdots$, generate a r.v. $\xi_{n+1|n}^i \sim Q_{n+1}(\xi_n^i, \cdot)$.

(iv) Stopping rule : define the stopping time

$$N_{n+1} = \inf\{N \, : \, \delta_{n+1}^2 \sum_{i=1}^{N} \Psi_{n+1}(\xi_{n+1|n}^i) \geq \sup_{x \in E} \Psi_{n+1}(x)\} \ ,$$

and set

$$\mu_{n+1|n}^{N_{n+1}} = S^{N_{n+1}}(Q_{n+1}\, \mu_n^{N_n}) = \frac{1}{N_{n+1}} \sum_{i=1}^{N_{n+1}} \delta_{\xi_{n+1|n}^i} \ .$$

Exactly as for the IPF, the resampling step (ii) requires to generate random variables according to the weighted discrete probability distribution $\mu_n^{N_n}$, which can be easily implemented. Notice that a.s.

$$\frac{1}{N} \sum_{i=1}^{N} \Psi_n(\xi_{n|n-1}^i) \longrightarrow \langle Q_n\, \mu_{n-1}^{N_{n-1}}, \Psi_n \rangle$$

as $N \uparrow \infty$, and if the nonnegative operator $R_n$ is mixing, then $\langle Q_n\, \mu_{n-1}^{N_{n-1}}, \Psi_n \rangle > 0$ in view of Remark 5.6, hence the stopping time $N_n$ is a.s. finite. Moreover, the normalization constant $c_n$ is positive, since

$$c_n = \sum_{i=1}^{N_n} \Psi_n(\xi_{n|n-1}^i) \geq \frac{1}{\delta_n^2} \sup_{x \in E} \Psi_n(x) > 0 \ .$$

First we check that the SPF satisfies (13). Indeed, for any $n \geq k \geq 1$ and for any bounded measurable function $F$ defined on $\mathcal{P}(E)$

$$\mathbb{E}[F(\mu_k^{N_k}) \mid Y_{1:n}] = \mathbb{E}[F(\mu_k^{N_k}) \mid Y_{1:k}] \ .$$

The following theorem shows that using a random number of particles allows to control the local error independently of any lower bound assumption on the likelihood functions. The counterpart is that the computational time of the resulting algorithm is random, and that the expected number of particles does depend on the integrated lower bounds of the likelihood functions.

**Theorem 5.12.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing, and the random number $N_k$ of particles is defined as in* (24)*, then the following inequality holds*

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, | \langle \mu_n - \mu_n^{N_n}, \phi \rangle | \mid Y_{1:n}\,] \leq \delta_n^{\rm W} + 2\, \frac{\delta_{n-1}^{\rm W}}{\varepsilon_n^2} + \frac{4}{\log 3} \sum_{k=1}^{n-2} \tau_{n:k+3}\, \frac{\delta_k^{\rm W}}{\varepsilon_{k+2}^2\, \varepsilon_{k+1}^2} \ ,$$

*where for any $k \geq 1$*

$$\delta_k^{\rm W} \leq 2\, \delta_k \sqrt{1 + \delta_k^2} \ ,$$

*and*

$$\frac{\rho_k}{\delta_k^2} \leq \mathbb{E}[N_k \mid Y_{1:k}] \leq \frac{\rho_k}{\delta_k^2}\, (1 + \delta_k^2) \ .$$

**Corollary 5.13.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing with $\varepsilon_k \geq \varepsilon > 0$, and the random number $N_k$ of particles is defined as in (24) with $\delta_k \leq \delta$, then convergence holds uniformly in time, i.e.*

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n - \mu_n^{N_n}, \phi \rangle\,|\,|\,Y_{1:n}\,] \leq (\,1 + \frac{2}{\varepsilon^2} + \frac{4}{\varepsilon^6\,\log 3}\,)\; 2\,\delta\,\sqrt{1+\delta^2}\ .$$

Proof of Theorem 5.12. It is sufficient to bound the local error $\delta_k^{\rm W}$ in the weak sense, and to apply Theorem 4.8. Since $R_k$ is mixing, $\langle Q_k\,\mu_{k-1}^{N_{k-1}}, \Psi_k \rangle > 0$ in view of Remark 5.6. Using estimate (22) with $\Lambda = \Psi_k$, $\mu = Q_k\,\mu_{k-1}^{N_{k-1}}$ and $\mathcal{F} = \sigma(Y_{1:k}\,,\,\mu_{k-1}^{N_{k-1}})$ yields

$$\begin{aligned} & \mathbb{E}[\,|\,\langle \mu_k^{N_k} - \bar{R}_k(\mu_{k-1}^{N_{k-1}}), \phi \rangle\,|\,|\,Y_{1:k}\,,\,\mu_{k-1}^{N_{k-1}}\,] \\ = \;& \mathbb{E}[\,|\,\langle \Psi_k \cdot S^{N_k}(Q_k\,\mu_{k-1}^{N_{k-1}}) - \Psi_k \cdot (Q_k\,\mu_{k-1}^{N_{k-1}}), \phi \rangle\,|\,|\,Y_{1:k}\,,\,\mu_{k-1}^{N_{k-1}}\,] \\ \leq \;& 2\,\delta_k\,\sqrt{1+\delta_k^2}\,\|\phi\|\ . \quad \Box \end{aligned}$$

In this section, we have proved that the IPF and its sequential variant converge uniformly in time under the mixing assumption. This theoretical argument is also supported by numerical evidence, e.g. in extreme cases where the hidden state sequence satisfies a noise–free state equation. Indeed, because multiple copies are produced after each resampling step, the diversity of the particle system can only decrease along the time in such cases, and the particle system ultimately concentrates on a few points, if not a single point, of the state space. This phenomenon is called *degeneracy of particle locations* and is another known cause of divergence of particle filters. To solve this degeneracy problem, and also the problem of *degeneracy of particle weights* already mentionned, we have proposed in Musso and Oudjane [23] to add a regularization step in the algorithm, so as to guarantee the diversity of the particle system along the time : the resulting filters, called regularized particle filters (RPF), are studied in the next section under the same mixing assumption.

# 6 Uniform convergence of regularized particle filters

The main idea consists in changing the discrete approximation $\mu_n^N$ for an absolutely continuous approximation, with the effect that in the resampling step $N$ random variables are generated according to an absolutely continuous distribution, hence producing a new particle system with $N$ different particle locations. In doing this, we implicitly assume that the hidden state sequence takes values in a Euclidean space $E = \mathbb{R}^m$, and that the optimal filter $\mu_n$ has a smooth density w.r.t. the Lebesgue measure, which is the case in most applications. From the theoretical point of view, this additional assumption allows to obtain strong approximations of the optimal filter, in total variation or in the $L^p$ sense for any $p \geq 1$. In practice, this provides approximate filters which are much more stable along the time than the IPF.

To obtain an absolutely continuous approximation is achieved by adding a regularization step in the algorithm, using a kernel method, classical in density estimation. If the regularization occurs before the correction by the likelihood function, we obtain the pre–regularized particle filter, the numerical analysis of which has been done in Le Gland, Musso and Oudjane [22], in the general case without the mixing assumption. An improved version of the pre–regularized particle filter, called the kernel filter, is proposed in Hürzeler and Künsch [18]. If the regularization occurs after the correction by the likelihood function, we obtain the post–regularized particle filter, which has been proposed in Musso and Oudjane [23] and in Oudjane and Musso [25] and compared with the IPF in some classical tracking problems, such as bearing–only tracking, or range and bearing tracking with multiple dynamical model.

The following notations and definitions will be used below. For any $\mu \in \mathcal{P}(E)$, define

$$I(\mu) := [\int_E |x|^{m+1}\,\mu(dx)\,]^{1/m+1}\ ,$$

and if $\mu$ is absolutely continuous w.r.t. the Lebesgue measure on $E$, with density $f = \dfrac{d\mu}{dx}$, define

$$I(f) = I(\mu) = [\int_E |x|^{m+1}\,f(x)\,dx\,]^{1/m+1} \qquad \text{and} \qquad J(f) := J(\frac{d\mu}{dx}) = \int_E \sqrt{f(x)}\,dx\ .$$

From the multidimensional Carlson inequality, see Lemma 7 in Holmström and Klemelä [16], there exists a universal constant $A_m$ such that, for any absolutely continuous $\mu \in \mathcal{P}(E)$

$$J(\frac{d\mu}{dx}) \leq A_m \, (I(\mu))^{m/2} \; , \tag{25}$$

hence $J(\dfrac{d\mu}{dx})$ is finite if $I(\mu)$ is finite.

Let the regularization kernel $K$ be a symmetric probability density on $E = \mathbb{R}^m$, such that

$$\int_E K(x)\, dx = 1 \; , \qquad \int_E x\, K(x)\, dx = 0 \qquad \text{and} \qquad \alpha := \tfrac{1}{2} \int_E |x|^2 \, K(x)\, dx < \infty \; .$$

Assume also that the regularization kernel $K$ is square–integrable, i.e.

$$\beta := [\int_E K^2(x)\, dx\,]^{1/2} < \infty$$

and that the symmetric probability density $L := \dfrac{K^2}{\beta^2}$ satisfies

$$\gamma := I(L) = [\int_E |x|^{m+1} \, L(x)\, dx\,]^{1/m+1} < \infty \; .$$

For any bandwidth $h > 0$, define the rescaled kernel

$$K_h(x) = \frac{1}{h^m} \, K(\frac{x}{h}) \, ,$$

for any $x \in E$.

**Definition 6.1.** *For any $\mu \in \mathcal{M}^+(E)$, the nonnegative measure $K_h * \mu$ is absolutely continuous w.r.t. the Lebesgue measure, with density*

$$\frac{d(K_h * \mu)}{dx}(x) = \int_E K_h(x - x') \, \mu(dx') \; ,$$

*where $*$ denotes the convolution operator.*

Notice that convolution by $K_h$ preserves the total mass, i.e. $K_h * \mu(E) = \mu(E)$ for any $\mu \in \mathcal{M}^+(E)$, hence $K_h * \bar{\mu}$ is the normalized nonnegative measure (i.e. probability distribution) associated with the unnormalized nonnegative measure $K_h * \mu$. Moreover, $K_h * \mu$ approximates $\mu$ in the following sense.

**Lemma 6.2.** *Let $\mu \in \mathcal{M}^+(E)$ be absolutely continuous w.r.t. the Lebesgue measure, with density $\dfrac{d\mu}{dx} \in W^{2,1}$. Then*

$$\|K_h * \mu - \mu\| \leq \alpha \, h^2 \, |\frac{d\mu}{dx}|_{2,1} \; .$$

The proof for the one dimensional case can be found in Silverman [27] or Devroye [13], and for the multidimensional case in Raviart [26] or Holmström and Klemelä [16].

Given a sample $(\xi^1, \cdots, \xi^N)$ from an unknown probability distribution $\mu \in \mathcal{P}(E)$ with a smooth density, and given a positive function $\Lambda$ which can be evaluated at any point of $E$, we are interested in approximating the projective product $\Lambda \cdot \mu$ by a probability distribution with a smooth density. By construction, the approximation error can be estimated in a strong sense such as the total variation. Of course, more sample points are needed to approximate a whole density, than are needed to simply approximate moments, as in the weak sense approximation. Typically, the sample size depends on the dimension, which is the usual *curse of dimensionality*. However, getting an approximation of the whole density is usually worth the effort, as it allows to get meaningful information, e.g. confidence regions.

The classical density estimation theory, see e.g. [27] for the $L^2$ theory and [13] for the $L^1$ theory, has extensively studied the problem of estimating $\mu$ alone, which reduces to our problem in the particular case where $\Lambda$ is constant. The solution consists in regularizing the empirical probability distribution associated with the sample and provides kernel–type estimators $K_h * S^N(\mu)$. Minimization of the mean errors $\mathbb{E}\|K_h * S^N(\mu) - \mu\|$ or $\mathbb{E}\|K_h * S^N(\mu) - \mu\|_2^2$, in the $L^1$ or $L^2$ sense, over the bandwidth $h$ and the regularization kernel $K$ has also been studied. Similarly, in our more general setting, we propose two kinds of estimators for $\Lambda \cdot \mu$ : the pre–regularized estimator $\Lambda \cdot (K_h * S^N(\mu))$ where the regularization occurs before the correction by $\Lambda$, and the post–regularized estimator $K_h * (\Lambda \cdot S^N(\mu))$ where the regularization occurs after the correction by $\Lambda$. We immediately see that the pre–regularized estimator consists in applying the correction by $\Lambda$ to the classical density estimator $K_h * S^N(\mu)$ and that both estimators reduce to the classical density estimator when $\Lambda$ is constant. Consequently, we will focus below on estimating the mean error for the post–regularized estimator, and results for the pre–regularized estimator will follow immediately. In Proposition 6.3, we consider the mean error between the unnormalized nonnegative measures $K_h * (\Lambda\, S^N(\mu))$ and $\Lambda\, \mu$. In the general case, the error between the corresponding probability distributions $K_h * (\Lambda \cdot S^N(\mu))$ and $\Lambda \cdot \mu$ will then be derived using estimate (4), but in some particular cases we may derive some sharper bounds. In what follows, we only state some bounds without trying to optimize over the bandwidth $h$ or the regularization kernel $K$.

**Proposition 6.3.** *Let $\mu \in \mathcal{P}(E)$ be absolutely continuous w.r.t. the Lebesgue measure, with density $\dfrac{d\mu}{dx} \in W^{2,1}$, and let $\Lambda$ be a nonnegative bounded measurable function defined on $E$, with bounded derivatives up to order two. Then*

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}|\langle K_h * (\Lambda\, S^N(\mu)) - \Lambda\, \mu, \phi\rangle| \leq \frac{1}{\sqrt{N}} \sup_{x\in E} \Lambda(x) + \alpha\, h^2\, |\Lambda\, \frac{d\mu}{dx}|_{2,1} \; .$$

*If in addition $I(\mu)$ is finite, then*

$$\mathbb{E}\|K_h * (\Lambda\, S^N(\mu)) - \Lambda\, \mu\| \leq \frac{\beta\, A_m}{\sqrt{N h^m}}\, (I(\mu) + h\, \gamma)^{m/2} \sup_{x\in E} \Lambda(x) + \alpha\, h^2\, |\Lambda\, \frac{d\mu}{dx}|_{2,1} \; .$$

The proof is based on the following decomposition of the error into variation and bias errors

$$K_h * (\Lambda\, S^N(\mu)) - \Lambda\, \mu = K_h * (\Lambda\, S^N(\mu)) - K_h * (\Lambda\, \mu) + K_h * (\Lambda\, \mu) - \Lambda\, \mu \; .$$

Under the assumptions, the nonnegative measure $\Lambda\, \mu$ is absolutely continuous w.r.t. the Lebesgue measure, with density $\Lambda\, \dfrac{d\mu}{dx} \in W^{2,1}$, such that

$$|\Lambda\, \frac{d\mu}{dx}|_{2,1} \leq \sup_{u\in W^{2,1}} \frac{|\Lambda\, u|_{2,1}}{\|u\|_{2,1}}\, \|\frac{d\mu}{dx}\|_{2,1} \; ,$$

and Lemma 6.2 can be used to bound the bias error. The following lemma is used to bound the variation error.

**Lemma 6.4.** *Let $\mu \in \mathcal{P}(E)$, and let $\Lambda$ be a nonnegative bounded measurable function defined on $E$. Then*

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}|\langle K_h * (\Lambda\, S^N(\mu)) - K_h * (\Lambda\, \mu), \phi\rangle| \leq \frac{1}{\sqrt{N}} \sup_{x\in E} \Lambda(x) \; .$$

*If in addition $I(\mu)$ is finite, then*

$$\mathbb{E}\|K_h * (\Lambda\, S^N(\mu)) - K_h * (\Lambda\, \mu)\| \leq \frac{\beta\, A_m}{\sqrt{N h^m}}\, (I(\mu) + h\, \gamma)^{m/2} \sup_{x\in E} \Lambda(x) \; .$$

Proof of Lemma 6.4. Using Lemma 5.1 yields

$$\begin{aligned} \mathbb{E}|\langle K_h * (\Lambda\, S^N(\mu)) - K_h * (\Lambda\, \mu), \phi\rangle\,| &= \mathbb{E}|\langle S^N(\mu) - \mu, \Lambda\,(K_h * \phi)\rangle\,| \\ &\leq \frac{1}{\sqrt{N}}\, \|\Lambda\,(K_h * \phi)\| \leq \frac{1}{\sqrt{N}}\, \|\phi\| \sup_{x\in E} \Lambda(x) \; , \end{aligned}$$

for any bounded measurable test function $\phi$ defined on $E$, which proves the estimate in the weak sense.

The proof of the estimate in total variation is classical. By definition

$$\|K_h * (\Lambda\, S^N(\mu)) - K_h * (\Lambda\, \mu)\| = \int_E |f_h^N(x) - f_h(x)|\, dx \; ,$$

where

$$f_h^N(x) = \frac{d(K_h * (\Lambda\, S^N(\mu)))}{dx}(x) = \frac{1}{N} \sum_{i=1}^N K_h(x-\xi^i)\, \Lambda(\xi^i) \;,$$

$$f_h(x) = \frac{d(K_h * (\Lambda\, \mu))}{dx}(x) = \int_E K_h(x-x')\, \Lambda(x')\, \mu(dx') \;,$$

and it follows from the proof of Lemma 5.1 that

$$\begin{aligned} \mathbb{E}|f_h^N(x) - f_h(x)| &\leq \frac{1}{\sqrt{N}} \, [ \int_E K_h^2(x-x')\, \Lambda^2(x')\, \mu(dx') \, ]^{1/2} \\ &\leq \frac{\beta}{\sqrt{N h^m}} \, [ \int_E L_h(x-x')\, \mu(dx') \, ]^{1/2} \; \sup_{x \in E} \Lambda(x) \;, \end{aligned}$$

where the symmetric probability density $L = \dfrac{K^2}{\beta^2}$ satisfies $K_h^2 = \dfrac{\beta^2}{h^m}\, L_h$. Therefore

$$\mathbb{E}\, \|K_h * (\Lambda\, S^N(\mu)) - K_h * (\Lambda\, \mu)\| \leq \frac{\beta}{\sqrt{N h^m}} \, J(\frac{d(L_h * \mu)}{dx}) \; \sup_{x \in E} \Lambda(x) \;.$$

Using the Minkowski inequality yields

$$\begin{aligned} I(L_h * \mu) &= [ \int_E \int_E |x' + h\, u|^{m+1}\, L(u)\, \mu(dx')\, du \, ]^{1/m+1} \\ &\leq [ \int_E |x'|^{m+1}\, \mu(dx') \, ]^{1/m+1} + h\, [ \int_E |u|^{m+1}\, L(u)\, du \, ]^{1/m+1} \\ &\leq I(\mu) + h\, I(L) \;, \end{aligned}$$

and using estimate (25) yields

$$J(\frac{d(L_h * \mu)}{dx}) \leq A_m \, (I(L_h * \mu))^{m/2} \leq A_m \, (I(\mu) + h\, \gamma)^{m/2} \;. \quad \square$$

**Remark 6.5.** If the probability distribution $\mu$ is absolutely continuous w.r.t. the Lebesgue measure, with probability density $f$, then the following more precise bound holds

$$J(\frac{d(L_h * \mu)}{dx}) = J(L_h * f) \leq J(f) + J(|L_h * f - f|) \;,$$

where $J(|L_h * f - f|)$ goes to zero when $h \downarrow 0$, see Proposition 8 in Holmström and Klemelä [16].

**Remark 6.6.** If in addition $\phi$, $\Lambda$ and $\mu$ are $\mathcal{F}$–measurable r.v.'s, and if conditionally w.r.t. $\mathcal{F}$ the r.v.'s $(\xi^1, \cdots, \xi^N)$ are i.i.d. with (conditional) probability distribution $\mu$, then the same estimates hold for conditional expectation w.r.t. $\mathcal{F}$, i.e.

$$\mathbb{E}[\, |\langle K_h * (\Lambda\, S^N(\mu)) - \Lambda\, \mu, \phi \rangle| \mid \mathcal{F}] \leq [\frac{1}{\sqrt{N}} \; \sup_{x \in E} \Lambda(x) + \alpha\, h^2\, |\Lambda\, \frac{d\mu}{dx}|_{2,1}] \; \|\phi\| \;, \tag{26}$$

and

$$\mathbb{E}[\, \|K_h * (\Lambda\, S^N(\mu)) - \Lambda\, \mu\| \mid \mathcal{F}] \leq \frac{\beta\, A_m}{\sqrt{N h^m}} \; (I(\mu) + h\, \gamma)^{m/2} \; \sup_{x \in E} \Lambda(x) + \alpha\, h^2\, |\Lambda\, \frac{d\mu}{dx}|_{2,1} \;. \tag{27}$$

### □ Pre–regularized particle filter

Let $\mu_n^{N,h}$ denote the pre–regularized particle filter (pre–RPF) approximation of $\mu_n$. The transition from $\mu_{n-1}^N$ to $\mu_n^{N,h}$ is described by the following diagram

$$\mu_{n-1}^{N,h} \xrightarrow[\text{prediction}]{\text{sampled}} \mu_{n|n-1}^{N,h} = S^N(Q_n\,\mu_{n-1}^{N,h}) \xrightarrow[\text{correction}]{\text{pre–regularized}} \mu_n^{N,h} = \Psi_n \cdot (K_h * \mu_{n|n-1}^{N,h}) \; .$$

In practice, the particle approximation $\mu_{n|n-1}^{N,h}$ is completely characterized by the particle system $\{\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^N\}$, and the transition from $\{\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^N\}$ to $\{\xi_{n+1|n}^1, \cdots, \xi_{n+1|n}^N\}$ consists of the following three steps.

(i) Correction : set

$$\mu_n^{N,h}(dx) = (\Psi_n \cdot (K_h * \mu_{n|n-1}^{N,h}))(dx) = \frac{1}{c_n} \sum_{i=1}^N \Psi_n(x)\,K_h(x - \xi_{n|n-1}^i)\;dx \; ,$$

with the normalization constant

$$c_n = \sum_{i=1}^N \int_E \Psi_n(x)\,K_h(x - \xi_{n|n-1}^i)\;dx = \sum_{i=1}^N \int_E \Psi_n(\xi_{n|n-1}^i + h\,u)\,K(u)\;du \; .$$

(ii) Resampling : independently for all $i = 1, \cdots, N$, generate a r.v. $\xi_n^i \sim \mu_n^{N,h}$.

(iii) Prediction : independently for all $i = 1, \cdots, N$, generate $\xi_{n+1|n}^i \sim Q_{n+1}(\xi_n^i, \cdot)$. Then set

$$\mu_{n+1|n}^{N,h} = S^N(Q_{n+1}\,\mu_n^{N,h}) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{n+1|n}^i} \; .$$

The resampling step (ii) requires to generate random variables according to a complex probability distribution known up to a normalization constant, which can be done with a rejection algorithm, see Devroye [12], or with the more efficient local rejection algorithm, see Hürzeler and Künsch [18], if the kernel $K$ has compact support. In any case, the implementation is less straightforward than for the IPF.

First we check that the pre–RPF satisfies (13). Indeed, for any $n \geq k \geq 1$ and for any bounded measurable function $F$ defined on $\mathcal{P}(E)$

$$\mathbb{E}[F(\mu_k^{N,h}) \mid Y_{1:n}] = \mathbb{E}[F(\mu_k^{N,h}) \mid Y_{1:k}] \; .$$

**Assumption R** (regularity of the Markov kernel) **:** For any $\mu \in \mathcal{P}(E)$, the probability distribution $Q_k\,\mu$ is absolutely continuous w.r.t. the Lebesgue measure, with density in $W^{2,1}$, and

$$D_k = \sup_{\mu \in \mathcal{P}(E)} |\frac{d(Q_k\,\mu)}{dx}|_{2,1} < \infty \; .$$

**Remark 6.7.** Assumption R is equivalent to suppose that for any $x \in E$, the probability distribution $Q_k(x, \cdot)$ is absolutely continuous w.r.t. the Lebesgue measure, with density $q_k(x, \cdot)$ in $W^{2,1}$ such that

$$D_k = \sup_{x \in E} |q_k(x, \cdot)|_{2,1} < \infty \; .$$

**Assumption M** (existence of moments) **:**

$$I_k = \sup_{\mu \in \mathcal{P}(E)} I(Q_k\,\mu) < \infty \; .$$

**Remark 6.8.** Assumption M is equivalent to suppose that

$$I_k = \sup_{x \in E} [\int_E |x'|^{m+1}\, Q_k(x, dx')\,]^{1/m+1} < \infty \; .$$

Alternatively, if the Markov kernel $Q_k$ is mixing, and $I(\mu_{k|k-1})$ is finite, then

$$I_k \leq \frac{1}{\varepsilon_k^{2/m+1}}\, I(\mu_{k|k-1}) < \infty \; .$$

**Theorem 6.9.** *If for any $k \geq 1$, Assumptions L and R hold, and the nonnegative operator $R_k$ is mixing, then the pre–RPF estimator satisfies*

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n - \mu_n^{N,h}, \phi \rangle\,|\,|\,Y_{1:n}\,] \leq \delta_n^{\mathrm{W}} + 2\,\frac{\delta_{n-1}^{\mathrm{W}}}{\varepsilon_n^2} + \frac{4}{\log 3} \sum_{k=1}^{n-2} \tau_{n:k+3}\, \frac{\delta_k^{\mathrm{W}}}{\varepsilon_{k+2}^2\, \varepsilon_{k+1}^2} \; ,$$

*where for any $k \geq 1$*

$$\delta_k^{\mathrm{W}} \leq [\,\frac{1}{\sqrt{N}} + \alpha\, h^2\, D_k\,]\; 2\, \rho_k \; .$$

*If in addition for any $k \geq 1$, Assumption M holds, then*

$$\mathbb{E}[\,\|\mu_n - \mu_n^{N,h}\|\,|\,Y_{1:n}\,] \leq \delta_n^{\mathrm{TV}} + \frac{2}{\log 3} \sum_{k=1}^{n-1} \tau_{n:k+2}\, \frac{\delta_k^{\mathrm{TV}}}{\varepsilon_{k+1}^2} \; ,$$

*where for any $k \geq 1$*

$$\delta_k^{\mathrm{TV}} \leq [\,\frac{\beta\, A_m}{\sqrt{N h^m}}\, (I_k + h\,\gamma)^{m/2} + \alpha\, h^2\, D_k\,]\; 2\, \rho_k \; .$$

The convergence result stated in Theorem 6.9 would still hold with a time dependent bandwidth, and with a time dependent number of particles.

PROOF OF THEOREM 6.9. To prove the estimate in the weak sense, it is sufficient to bound the local error $\delta_k^{\mathrm{W}}$ in the weak sense, and to apply Theorem 4.8. Using estimate (3) yields

$$\begin{aligned}
|\,\langle \mu_k^{N,h} - \bar{R}_k(\mu_{k-1}^{N,h}), \phi \rangle\,| \;&=\; |\,\langle \Psi_k \cdot (K_h * S^N(Q_k\, \mu_{k-1}^{N,h})) - \Psi_k \cdot (Q_k\, \mu_{k-1}^{N,h}), \phi \rangle\,| \\
&\leq\; \frac{|\,\langle K_h * S^N(Q_k\, \mu_{k-1}^{N,h}) - Q_k\, \mu_{k-1}^{N,h}, \Psi_k\, \phi \rangle\,|}{\langle Q_k\, \mu_{k-1}^{N,h}, \Psi_k \,\rangle} \\
&\quad + \frac{|\,\langle K_h * S^N(Q_k\, \mu_{k-1}^{N,h}) - Q_k\, \mu_{k-1}^{N,h}, \Psi_k \rangle\,|}{\langle Q_k\, \mu_{k-1}^{N,h}, \Psi_k \,\rangle}\; \|\phi\| \; ,
\end{aligned}$$

for any bounded measurable test function $\phi$ defined on $E$. By definition

$$\langle Q_k\, \mu_{k-1}^{N,h}, \Psi_k \rangle \geq \inf_{\mu \in \mathcal{P}(E)} \langle Q_k\, \mu, \Psi_k \rangle \; .$$

Using estimate (26) with $\Lambda \equiv 1$, $\mu = Q_k\, \mu_{k-1}^{N,h}$ and $\mathcal{F} = \sigma(Y_{1:k}\,,\, \mu_{k-1}^{N,h})$, yields

$$\begin{aligned}
&\mathbb{E}[\,|\langle K_h * S^N(Q_k\, \mu_{k-1}^{N,h}) - Q_k\, \mu_{k-1}^{N,h}, \Psi_k\, \phi \rangle\,|\,|\,Y_{1:k}\,,\, \mu_{k-1}^{N,h}] \\
&\leq\; [\frac{1}{\sqrt{N}} + \alpha\, h^2\, |\frac{d(Q_k\, \mu_{k-1}^{N,h})}{dx}|_{2,1}]\; \|\Psi_k\, \phi\| \\
&\leq\; [\frac{1}{\sqrt{N}} + \alpha\, h^2\, D_k]\; \|\phi\|\; \sup_{x \in E} \Psi_k(x) \; .
\end{aligned}$$

To prove the estimate in total variation, it is sufficient to bound the local error $\delta_k^{\rm TV}$ in total variation, and to apply Theorem 4.6. Using estimate (4) yields

$$\begin{aligned} \|\mu_k^{N,h} - \bar{R}_k(\mu_{k-1}^{N,h})\| &= \|\Psi_k \cdot (K_h * S^N(Q_k\,\mu_{k-1}^{N,h})) - \Psi_k \cdot (Q_k\,\mu_{k-1}^{N,h})\| \\ &\leq 2\,\frac{\sup_{x \in E} \Psi_k(x)}{\langle Q_k\,\mu_{k-1}^{N,h}, \Psi_k\,\rangle}\; \|K_h * S^N(Q_k\,\mu_{k-1}^{N,h}) - Q_k\,\mu_{k-1}^{N,h}\| \; . \end{aligned}$$

Using estimate (27) with $\Lambda \equiv 1$, $\mu = Q_k\,\mu_{k-1}^{N,h}$ and $\mathcal{F} = \sigma(Y_{1:k}\;,\;\mu_{k-1}^{N,h})$, yields

$$\begin{aligned} &\mathbb{E}[\,\|K_h * S^N(Q_k\,\mu_{k-1}^{N,h}) - Q_k\,\mu_{k-1}^{N,h}\,\|\mid Y_{1:k}\;,\;\mu_{k-1}^{N,h}] \\ &\leq \frac{\beta\,A_m}{\sqrt{N h^m}}\,(I(Q_k\,\mu_{k-1}^{N,h}) + h\,\gamma)^{m/2} + \alpha\,h^2\,|\frac{d(Q_k\,\mu_{k-1}^{N,h})}{dx}|_{2,1} \\ &\leq \frac{\beta\,A_m}{\sqrt{N h^m}}\,(I_k + h\,\gamma)^{m/2} + \alpha\,h^2\,D_k\;. \quad \Box \end{aligned}$$

In the pre–RPF, the correction is applied directly to a regularized probability distribution, hence each point in the support of the regularized density is updated, and in principle the *degeneracy of particle weights* which occurs when the correction is applied to a discrete probability distribution, as in the IPF, is now avoided. This intuition is supported by numerical evidence, and by the following theorem, which shows that it is possible to control the local error, averaged over observation sequences, independently of any lower bound assumption on the likelihood functions (notice that the a.s. bounds of Theorem 6.9 still depend on the integrated lower bounds of the likelihood functions).

**Theorem 6.10.** *If for any $k \geq 1$, Assumptions R and M hold, and the nonnegative operator $R_k$ is mixing with nonrandom $\varepsilon_k$, then the pre–RPF estimator satisfies*

$$\mathbb{E}\|\mu_n - \mu_n^{N,h}\| \leq \delta_n + \frac{2}{\log 3}\,\sum_{k=1}^{n-1} \tau_{n:k+2}\,\frac{\delta_k}{\varepsilon_{k+1}^2}\;,$$

*where for any $k \geq 1$*

$$\delta_k \leq \frac{2}{\varepsilon_k^2}\,[\,\frac{\beta\,A_m}{\sqrt{N h^m}}\,(I_k + h\,\gamma)^{m/2} + \alpha\,h^2\,D_k\,]\;.$$

**Corollary 6.11.** *If for any $k \geq 1$, the nonnegative operator $R_k$ is mixing with $\varepsilon_k \geq \varepsilon > 0$ and nonrandom $\varepsilon$, and Assumptions R and M hold with $D_k \leq D$ and $I_k \leq I$, then convergence, averaged over observation sequences, holds uniformly in time, i.e.*

$$\mathbb{E}\|\mu_n - \mu_n^{N,h}\| \leq (\,1 + \frac{2}{\varepsilon^4\,\log 3}\,)\;\delta\;,$$

*with*

$$\delta \leq \frac{2}{\varepsilon^2}\,[\,\frac{\beta\,A_m}{\sqrt{N h^m}}\,(I + h\,\gamma)^{m/2} + \alpha\,h^2\,D\,]\;.$$

Both the SPF, see Theorem 5.12, and the pre–RPF allow to bound the error independently of any lower bound assumption on the likelihood functions, and the computational time of both algorithms is random (recall that a rejection is needed in the resampling step (ii) of the pre–RPF).

PROOF OF THEOREM 6.10. It is sufficient to bound the local error $\mathbb{E}[\delta_k^{\rm TV}]$ in total variation, averaged over observation sequences, and to apply Theorem 4.6. Using the mixing property of the nonnegative operator $R_k$ yields

$$\langle Q_k\,\mu_{k-1}^{N,h}, \Psi_k\rangle \geq \varepsilon_k^2\,\langle Q_k\,\mu_{k-1}, \Psi_k\rangle = \varepsilon_k^2\,\langle \mu_{k|k-1}, \Psi_k\rangle\;,$$

with nonrandom $\varepsilon_k$, hence using inequality (4) yields

$$\|\mu_k^{N,h} - \bar{R}_k(\mu_{k-1}^{N,h})\| \leq 2 \int_E \frac{\Psi_k(x)}{\langle Q_k\, \mu_{k-1}^{N,h}, \Psi_k \rangle} \, |\mu_k|(dx) \leq \frac{2}{\varepsilon_k^2} \int_E \frac{\Psi_k(x)}{\langle \mu_{k|k-1}, \Psi_k \rangle} \, |\mu_k|(dx) \; ,$$

where

$$\mu_k = K_h * S^N(Q_k\, \mu_{k-1}^{N,h}) - Q_k\, \mu_{k-1}^{N,h} \; .$$

It follows from Remark 2.2 that

$$\mathbb{E}[\, \frac{\Psi_k(x)}{\langle \mu_{k|k-1}, \Psi_k \rangle} \mid Y_{1:k-1} \, , \, \mu_{k|k-1}^{N,h} \, , \, \mu_{k-1}^{N,h} \,] = \mathbb{E}[\, \frac{\Psi_k(x)}{\langle \mu_{k|k-1}, \Psi_k \rangle} \mid Y_{1:k-1} \,] = 1 \; ,$$

hence

$$\mathbb{E}[\, \|\mu_k^{N,h} - \bar{R}_k(\mu_{k-1}^{N,h})\| \mid Y_{1:k-1} \, , \, \mu_{k|k-1}^{N,h} \, , \, \mu_{k-1}^{N,h} \,] \leq \frac{2}{\varepsilon_k^2} \, |\mu_k|(E) \; .$$

Using estimate (27) with $\Lambda \equiv 1$, $\mu = Q_k\, \mu_{k-1}^{N,h}$ and $\mathcal{F} = \sigma(\mu_{k-1}^{N,h})$, yields

$$\begin{aligned}
\mathbb{E}[\, |\mu_k|(E) \mid \mu_{k-1}^{N,h} \,] &= \mathbb{E}[\, \|K_h * S^N(Q_k\, \mu_{k-1}^{N,h}) - Q_k\, \mu_{k-1}^{N,h}\| \mid \mu_{k-1}^{N,h} \,] \\
&\leq \frac{\beta\, A_m}{\sqrt{N h^m}} \, (I(Q_k\, \mu_{k-1}^{N,h}) + h\, \gamma)^{m/2} + \alpha\, h^2 \, |\frac{d(Q_k\, \mu_{k-1}^{N,h})}{dx}|_{2,1} \\
&\leq \frac{\beta\, A_m}{\sqrt{N h^m}} \, (I_k + h\, \gamma)^{m/2} + \alpha\, h^2 \, D_k \; . \quad \Box
\end{aligned}$$

**Remark 6.12.** In the same way as for the SPF, see Theorem 5.12, one could think of using a random number of particles, so as to avoid any lower bound assumption on the likelihood functions. However, for the pre–RPF, it is not sufficient to evaluate the quantity $\Psi_k(\xi_{k|k-1}^i)$ for each simulated particle, but one has to evaluate the integral

$$\int_E \Psi_n(x)\, K_h(x - \xi_{n|n-1}^i)\, dx \; ,$$

instead. This evaluation is in general very costly, which makes the idea of using a random number of particles for the pre–RPF rather unpractical.

## □ Post–regularized particle filter

Let $\mu_n^{N,h}$ denote the post–regularized particle filter (post–RPF) approximation of $\mu_n$. The transition from $\mu_{n-1}^{N,h}$ to $\mu_n^{N,h}$ is described by the following diagram

$$\mu_{n-1}^{N,h} \xrightarrow[\text{prediction}]{\text{sampled}} \mu_{n|n-1}^{N,h} = S^N(Q_n\, \mu_{n-1}^{N,h}) \xrightarrow[\text{correction}]{\text{post–regularized}} \mu_n^{N,h} = K_h * (\Psi_n \cdot \mu_{n|n-1}^{N,h}) \; .$$

In practice, the particle approximation $\mu_{n|n-1}^{N,h}$ is completely characterized by the particle system $\{\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^N\}$, and the transition from $\{\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^N\}$ to $\{\xi_{n+1|n}^1, \cdots, \xi_{n+1|n}^N\}$ consists of the following three steps.

(i) Correction : for all $i = 1, \cdots, N$, compute the weight

$$\omega_n^i = \frac{1}{c_n} \, \Psi_n(\xi_{n|n-1}^i) \; ,$$

with the normalization constant

$$c_n = \sum_{i=1}^N \Psi_n(\xi_{n|n-1}^i) \; .$$

Then set

$$\mu_n^{N,h}(dx) = (K_h * (\Psi_n \cdot \mu_{n|n-1}^{N,h}))(dx) = \sum_{i=1}^N \omega_n^i \, K_h(x - \xi_{n|n-1}^i) \, dx \; .$$

(ii) Resampling : independently for all $i = 1, \cdots, N$, generate a r.v. $\xi_n^i \sim \mu_n^{N,h}$.

(iii) Prediction : independently for all $i = 1, \cdots, N$, generate a r.v. $\xi_{n+1|n}^i \sim Q_{n+1}(\xi_n^i, \cdot)$. Then set

$$\mu_{n+1|n}^{N,h} = S^N(Q_{n+1}\, \mu_n^{N,h}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_{n+1|n}^i} \; .$$

The resampling step (ii) requires to generate random variables according to the weighted mixture $\mu_n^{N,h}$ of rescaled kernels, which can be easily implemented.

First we check that the post–RPF satisfies (13). Indeed, for any $n \geq k \geq 1$ and for any bounded measurable function $F$ defined on $\mathcal{P}(E)$

$$\mathbb{E}[F(\mu_k^{N,h}) \mid Y_{1:n}] = \mathbb{E}[F(\mu_k^{N,h}) \mid Y_{1:k}] \; .$$

**Assumption L" :**

$$\sup_{u \in W^{2,1}} \frac{|\Psi_k\, u|_{2,1}}{\|u\|_{2,1}} < \infty \; .$$

**Remark 6.13.** If $\Psi_k$ is bounded, with bounded derivatives up to order two, then

$$\sup_{u \in W^{2,1}} \frac{|\Psi_k\, u|_{2,1}}{\|u\|_{2,1}} < \infty \; .$$

If Assumption L" holds, and if for any $k \geq 1$, the nonnegative operator $R_k$ is mixing, then the following notation is introduced

$$\rho_k'' := \frac{\displaystyle\sup_{u \in W^{2,1}} \frac{|\Psi_k\, u|_{2,1}}{\|u\|_{2,1}}}{\displaystyle\inf_{\mu \in \mathcal{P}(E)} \langle Q_k\, \mu, \Psi_k \rangle} \; ,$$

and in view of Remark 5.6, $\rho_k''$ is a.s. finite.

**Assumption R"** (additional regularity of the Markov kernel) **:** For any $\mu \in \mathcal{P}(E)$, the probability distribution $Q_k\, \mu$ is absolutely continuous w.r.t. the Lebesgue measure, with density in $W^{2,1}$, and

$$D_k'' = \sup_{\mu \in \mathcal{P}(E)} \|\frac{d(Q_k\, \mu)}{dx}\|_{2,1} < \infty \; .$$

**Remark 6.14.** Assumption R" is equivalent to suppose that for any $x \in E$, the probability distribution $Q_k(x, \cdot)$ is absolutely continuous w.r.t. the Lebesgue measure, with density $q_k(x, \cdot)$ in $W^{2,1}$ such that

$$D_k'' = \sup_{x \in E} \|q_k(x, \cdot)\|_{2,1} < \infty \; .$$

**Theorem 6.15.** *If for any $k \geq 1$, Assumptions L" and R" hold, and the nonnegative operator $R_k$ is mixing, then the post–RPF estimator satisfies*

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, |\, \langle \mu_n - \mu_n^{N,h}, \phi \rangle \,|\, \mid Y_{1:n}\,] \leq \delta_n^{\rm W} + 2\, \frac{\delta_{n-1}^{\rm W}}{\varepsilon_n^2} + \frac{4}{\log 3} \sum_{k=1}^{n-2} \tau_{n:k+3}\, \frac{\delta_k^{\rm W}}{\varepsilon_{k+2}^2\, \varepsilon_{k+1}^2} \; ,$$

*where for any $k \geq 1$*

$$\delta_k^{\rm W} \leq \frac{1}{\sqrt{N}}\, 2\, \rho_k + \alpha\, h^2\, D_k''\, \rho_k'' \; .$$

*If in addition for any $k \geq 1$, Assumption M holds, then*

$$\mathbb{E}[\, \|\mu_n - \mu_n^{N,h}\| \mid Y_{1:n}\,] \leq \delta_n^{\rm TV} + \frac{2}{\log 3} \sum_{k=1}^{n-1} \tau_{n:k+2}\, \frac{\delta_k^{\rm TV}}{\varepsilon_{k+1}^2} \; ,$$

*where for any $k \geq 1$*

$$\delta_k^{\rm TV} \leq [\frac{1}{\sqrt{N}} + \frac{\beta\, A_m}{\sqrt{N h^m}}\, (I_k + h\, \gamma)^{1/2}\,]\, \rho_k + \alpha\, h^2\, D_k''\, \rho_k'' \; .$$

PROOF OF THEOREM 6.15. The proof is similar to the proof of Theorem 6.9 except that estimates (26) and (27) are used here with $\Lambda = \Psi_k$.

To prove the estimate in the weak sense, it is sufficient to bound the local error $\delta_k^{\mathrm{W}}$ in the weak sense, and to apply Theorem 4.8. Since convolution by $K_h$ preserves the total mass, using estimate (3) yields

$$\begin{aligned}
|\langle \mu_k^{N,h} - \bar{R}_k(\mu_{k-1}^{N,h}), \phi \rangle| &= |\langle K_h * (\Psi_k \cdot S^N(Q_k\,\mu_{k-1}^{N,h})) - \Psi_k \cdot (Q_k\,\mu_{k-1}^{N,h}), \phi \rangle| \\
&\leq \frac{|\langle K_h * (\Psi_k\, S^N(Q_k\,\mu_{k-1}^{N,h})) - \Psi_k\,(Q_k\,\mu_{k-1}^{N,h}), \phi \rangle|}{\langle Q_k\,\mu_{k-1}^{N,h}, \Psi_k \rangle} \\
&\quad + \frac{|\langle S^N(Q_k\,\mu_{k-1}^{N,h}) - Q_k\,\mu_{k-1}^{N,h}, \Psi_k \rangle|}{\langle Q_k\,\mu_{k-1}^{N,h}, \Psi_k \rangle}\, \|\phi\| ,
\end{aligned}$$

for any bounded measurable test function $\phi$ defined on $E$. By definition

$$\langle Q_k\,\mu_{k-1}^{N,h}, \Psi_k \rangle \geq \inf_{\mu \in \mathcal{P}(E)} \langle Q_k\,\mu, \Psi_k \rangle .$$

Under Assumption R"

$$|\Psi_k \frac{d(Q_k\,\mu)}{dx}|_{2,1} \leq \|\frac{d(Q_k\,\mu)}{dx}\|_{2,1} \sup_{u \in W^{2,1}} \frac{|\Psi_k\,u|_{2,1}}{\|u\|_{2,1}} \leq D_k'' \sup_{u \in W^{2,1}} \frac{|\Psi_k\,u|_{2,1}}{\|u\|_{2,1}} .$$

Using estimate (26) with $\Lambda = \Psi_k$, $\mu = Q_k\,\mu_{k-1}^{N,h}$ and $\mathcal{F} = \sigma(Y_{1:k}\,,\,\mu_{k-1}^{N,h})$, yields

$$\begin{aligned}
&\mathbb{E}[\,|\langle K_h * (\Psi_k\, S^N(Q_k\,\mu_{k-1}^{N,h})) - \Psi_k\,(Q_k\,\mu_{k-1}^{N,h}), \phi \rangle|\mid Y_{1:k}\,,\,\mu_{k-1}^{N,h}] \\
&\leq [\frac{1}{\sqrt{N}} \sup_{x \in E} \Psi_k(x) + \tfrac{1}{2}\,\alpha\,h^2\,|\Psi_k \frac{d(Q_k\,\mu_{k-1}^{N,h})}{dx}|_{2,1}\,]\,\|\phi\| \\
&\leq [\frac{1}{\sqrt{N}} \sup_{x \in E} \Psi_k(x) + \tfrac{1}{2}\,\alpha\,h^2\,D_k'' \sup_{u \in W^{2,1}} \frac{|\Psi_k\,u|_{2,1}}{\|u\|_{2,1}}\,]\,\|\phi\| .
\end{aligned}$$

Using estimate (21) with $\phi = \Psi_k$, $\mu = Q_k\,\mu_{k-1}^{N,h}$ and $\mathcal{F} = \sigma(Y_{1:k}\,,\,\mu_{k-1}^{N,h})$, yields

$$\mathbb{E}[\,|\langle S^N(Q_k\,\mu_{k-1}^{N,h}) - Q_k\,\mu_{k-1}^{N,h}, \Psi_k \rangle|\mid Y_{1:k}\,,\,\mu_{k-1}^{N,h}] \leq \frac{1}{\sqrt{N}} \sup_{x \in E} \Psi_k(x) .$$

To prove the estimate in total variation, it is sufficient to bound the local error $\delta_k^{\mathrm{TV}}$ in total variation, and to apply Theorem 4.6. Since convolution by $K_h$ preserves the total mass, using estimate (4) yields

$$\begin{aligned}
\|\mu_k^{N,h} - \bar{R}_k(\mu_{k-1}^{N,h})\| &= \|K_h * (\Psi_k \cdot S^N(Q_k\,\mu_{k-1}^{N,h})) - \Psi_k \cdot (Q_k\,\mu_{k-1}^{N,h})\| \\
&\leq \frac{\|K_h * (\Psi_k\, S^N(Q_k\,\mu_{k-1}^{N,h})) - \Psi_k\,(Q_k\,\mu_{k-1}^{N,h})\|}{\langle Q_k\,\mu_{k-1}^{N,h}, \Psi_k \rangle} \\
&\quad + \frac{|\langle S^N(Q_k\,\mu_{k-1}^{N,h}) - Q_k\,\mu_{k-1}^{N,h}, \Psi_k \rangle|}{\langle Q_k\,\mu_{k-1}^{N,h}, \Psi_k \rangle} .
\end{aligned}$$

Using estimate (27) with $\Lambda = \Psi_k$, $\mu = Q_k\,\mu_{k-1}^{N,h}$ and $\mathcal{F} = \sigma(Y_{1:k}\,,\,\mu_{k-1}^{N,h})$, yields

$$\begin{aligned}
&\mathbb{E}[\,\|K_h * (\Psi_k\, S^N(Q_k\,\mu_{k-1}^{N,h})) - \Psi_k\,(Q_k\,\mu_{k-1}^{N,h})\|\mid Y_{1:k}\,,\,\mu_{k-1}^{N,h}] \\
&\leq \frac{\beta\,A_m}{\sqrt{N h^m}}\,(I(Q_k\,\mu_{k-1}^{N,h}) + h\,\gamma)^{m/2} \sup_{x \in E} \Psi_k(x) + \alpha\,h^2\,|\Psi_k \frac{d(Q_k\,\mu_{k-1}^{N,h})}{dx}|_{2,1} \\
&\leq \frac{\beta\,A_m}{\sqrt{N h^m}}\,(I_k + h\,\gamma)^{m/2} \sup_{x \in E} \Psi_k(x) + \alpha\,h^2\,D_k'' \sup_{u \in W^{2,1}} \frac{|\Psi_k\,u|_{2,1}}{\|u\|_{2,1}} . \quad \square
\end{aligned}$$

# References


[1] R. Atar. Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain. *The Annals of Probability*, 26(4):1552–1574, Oct. 1998.

[2] R. Atar, F. Viens, and O. Zeitouni. Robustness of Zakai's equation via Feynman–Kac representations. In W. M. McEneaney, G. G. Yin, and Q. Zhang, editors, *Stochastic Analysis, Control, Optimization and Applications : A Volume in Honor of W.H. Fleming*, Systems & Control : Foundations & Applications, pages 339–352. Birkäuser, Boston, 1999.

[3] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 33(6):697–725, 1997.

[4] R. Atar and O. Zeitouni. Lyapunov exponents for finite state nonlinear filtering. *SIAM Journal on Control and Optimization*, 35(1):36–55, Jan. 1997.

[5] G. Birkhoff. *Lattice Theory*, volume XXV of *Colloquium Publications*. AMS, Providence, RI, 3rd edition, 1967.

[6] A. S. Budhiraja and D. L. Ocone. Exponential stability of discrete–time filters for bounded observation noise. *Systems & Control Letters*, 30(4):185–193, 1997.

[7] A. S. Budhiraja and D. L. Ocone. Exponential stability in discrete–time filtering for non–ergodic signals. *Stochastic Processes and their Applications*, 82(2):245–257, 1999.

[8] G. Da Prato, M. Fuhrman, and P. Malliavin. Asymptotic ergodicity of the process of conditional law in some problem of nonlinear filtering. *Journal of Functional Analysis*, 164(2):356–377, 1999.

[9] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 37(2):155–194, 2001.

[10] P. Del Moral and J. Jacod. The Monte–Carlo method for filtering with discrete–time observations : Central limit theorems. Prépublication 515, Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris VI, July 1999. `http://www.proba.jussieu.fr/mathdoc/textes/PMA-515.dvi`.

[11] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering. In J. Azéma, M. Émery, M. Ledoux, and M. Yor, editors, *Séminaire de Probabilités XXXIV*, volume 1729 of *Lecture Notes in Mathematics*, pages 1–145. Springer–Verlag, Berlin, 2000.

[12] L. Devroye. *Non–Uniform Random Variate Generation*. Springer–Verlag, New York, 1986.

[13] L. Devroye. *A Course on Density Estimation*, volume 14 of *Progress in Probability and Statistics*. Birkhäuser, Boston, 1987.

[14] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer–Verlag, New York, 2001.

[15] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear / non–Gaussian Bayesian state estimation. *IEE Proceedings, Part F*, 140(2):107–113, Apr. 1993.

[16] L. Holmström and J. Klemelä. Asymptotic bounds for the expected $L_1$ error of a multivariate kernel density estimator. *Journal of Multivariate Analysis*, 42(2):245–266, 1992.

[17] E. Hopf. An inequality for positive integral linear operators. *Journal of Mathematics & Mechanics*, 12(5):683–692, 1963.

[18] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state space models. *Journal of Computational and Graphical Statistics*, 7(2):175–193, June 1998.

[19] H. Kunita. Ergodic properties of nonlinear filtering processes. In K. S. Alexander and J. C. Watkins, editors, *Spatial Stochastic Processes : Festschrift in honor of T.E. Harris*, volume 19 of *Progress in Probability*, pages 233–256. Birkhäuser, Boston, 1991.

[20] F. Le Gland and L. Mevel. Basic properties of the projective product, with application to products of column–allowable nonnegative matrices. *Mathematics of Control, Signals, and Systems*, 13(1):41–62, 2000.

[21] F. Le Gland and L. Mevel. Exponential forgetting and geometric ergodicity in hidden Markov models. *Mathematics of Control, Signals, and Systems*, 13(1):63–93, 2000.

[22] F. Le Gland, C. Musso, and N. Oudjane. An analysis of regularized interacting particle methods for nonlinear filtering. In J. Rojíček, M. Valečkova, M. Kárný, and K. Warwick, editors, *Preprints of the 3rd IEEE European Workshop on Computer–Intensive Methods in Control and Data Processing, Prague 1998*, pages 167–174, Sept. 1998.

[23] C. Musso and N. Oudjane. Regularization schemes for branching particle systems as a numerical solving method of the nonlinear filtering problem. In *Proceedings of the Irish Signals and Systems Conference, Dublin 1998*, June 1998.

[24] D. L. Ocone and E. Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM Journal on Control and Optimization*, 34(1):226–243, Jan. 1996.

[25] N. Oudjane and C. Musso. Multiple model particle filter. In *17ème Colloque GRETSI, Vannes 1999*, pages 681–684, Sept. 1999.

[26] P.-A. Raviart. An analysis of particle methods. In F. Brezzi, editor, *Numerical Methods in Fluid Dynamics, Como 1983*, volume 1127 of *Lecture Notes in Mathematics*, pages 243–324. Springer–Verlag, Berlin, 1985.

[27] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1986.

[28] Ł. Stettner. Invariant measures for the pair : state, approximate filtering process. *Colloquium Mathematicum*, LXII(2):347–351, 1991.

# Contents

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399